Carnegie Mellon
Software Engineering Institute
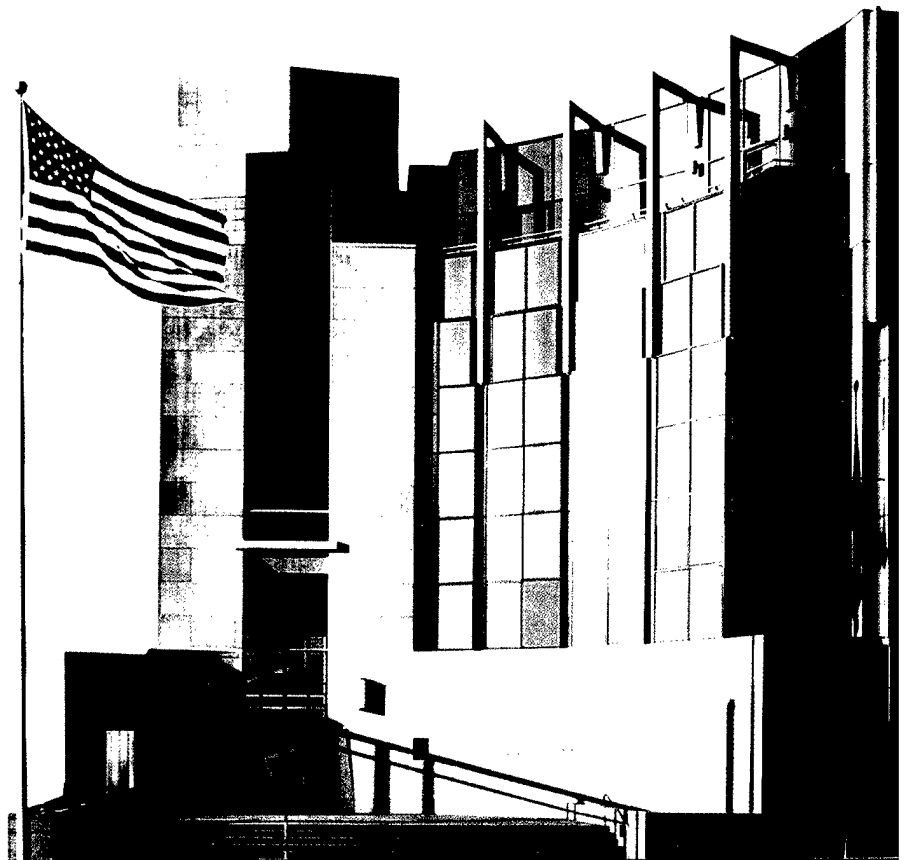
An Evaluation Theory Perspective of the Architecture Tradeoff Analysis Method^{SM} (ATAM^{SM})

Marta Lopez

September 2000

20010312 135

An Evaluation Theory Perspective of the Architecture Tradeoff Analysis Method^{SM} (ATAM^{SM})

Marta Lopez

September 2000

Product Line Systems
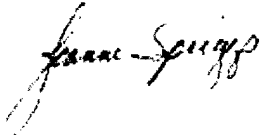
This report was prepared for the

The ideas and findings in this report should not be construed as an official DoD position. It is published in the interest of scientific and technical information exchange.

FOR THE COMMANDER

Joanne E. Spriggs
Contracting Office Representative

Table of Contents

List of Figures

List of Tables .

Abstract

Evaluation is a key analytical process in all disciplines and intellectual and practical endeavors. Also it is a key process in the software engineering field in which it is possible to apply different types of evaluation methods. The study of diverse evaluation methods performed in software and non-software disciplines and theoretical concepts could provide knowledge of the complexity and ubiquity of this important process. This study was the basis to obtain a set of basic evaluation components. They constitute a framework that can be used to developed a new evaluation method or review an existing one with the purpose of improving the development of the method being analyzed. In particular, this framework had been applied to review the Architecture Tradeoff Analysis Method[SM] (ATAM[SM]) by means of the identification of the evaluation components and the analysis of their development or elicitation. In this paper, the target, evaluation criteria, yardstick, data-gathering techniques, synthesis techniques and evaluation process of ATAM have been identified and analyzed. The most relevant conclusions are the role of stakeholders and the significance of attribute-based architectural styles (ABASs) in an ATAM evaluation.

[SM] Architectural Tradeoff Analysis Method and ATAM are service marks of Carnegie Mellon University.

1 Introduction

Evaluation is a key analytical process in all disciplines and intellectual and practical endeavors. Also it is a key process in the software engineering field in which it is possible to apply different types of evaluation methods. The study of diverse evaluation methods performed in other disciplines and theoretical concepts could provide knowledge of the complexity and ubiquity of this important process. In particular, we can obtain six essential components upon which an evaluation method could be developed: target, criteria, yardstick, data-gathering techniques, synthesis techniques, and evaluation process. These six components constitute a framework that we can use to develop a method of evaluation. This framework will be used to analyze the Architecture Tradeoff Analysis Method (ATAM) in order to find out whether it develops all the components of an evaluation. The final goal of this study is the analysis of the complete definition of the ATAM and the proposals or suggestions to improve the method description. With this aim in mind, the evaluation components of the framework are described in Section 2; in Section 3, current ATAM definitions will be matched with the framework; and finally, Section 4 presents a summary of the matching and the suggestions to improve the method.

2 Evaluation Concepts

In this section, we present the framework against which the ATAM will be analyzed. To develop this framework, both the Discipline of Evaluation and the Evaluation Theory were analyzed to determine the main components of an evaluation. A brief description of this study is included in Section 2.1. The framework and all evaluation components are described in Section 2.2. However, before covering this detail, it is necessary to understand the concept of "evaluation."

In general, an evaluation can be defined as the process of determining merit, worth, or significance. However, there are also common synonyms for the terms in this definition: "quality" is often used instead of "merit," "value" instead of "worth," and "importance" instead of "significance." Also, due to the fact that the evaluation is an anxiety-provoking activity for most people, other synonyms for the term can be found: analysis appraisal, audit, review, examination, and so on [Scriven 00]. In general, an evaluation involves the following [Scriven 91]:

- identification of relevant standards of merit, worth, or value
- investigation of the performance of targets (whatever is being evaluated) on these standards
- integration or synthesis of the results to achieve an overall evaluation result or a set of association evaluation results

Based on this, an evaluation can be differentiated from the simple information-gathering activity and the measurement process. The evaluators will have to collect data about the target in order to investigate its performance. So, information gathering is one task during the evaluation process, but it is not the entire evaluation. Also, in this context, measures are considered a technique through which we will obtain information about the target.

2.1 The Discipline and Theory of Evaluation

Evaluation can be described as an ubiquitous process because we can find it everywhere. Due to its use in many different disciplines, evaluation has been considered as a section of other disciplines instead of a discipline in itself. Nevertheless, according to Scriven, evaluation is

> *"one of the most powerful and versatile of the "transdisciplines" – tool disciplines such as logic, design, and statistics – that apply across broad ranges of the human investigative and creative effort while maintaining the autonomy of a discipline in their own right"* [Scriven 91].

From this perspective, the developed evaluation theories, methods, and lessons learned in different disciplines can be analyzed and later applied to improve evaluation practice in a given discipline.

At the present time there is *no general* Theory of Evaluation totally developed and common to all the disciplines and areas of knowledge in which the evaluation is applied. On the contrary, there are diverse theories of program evaluation, each one focused only on a specific method [Shadish 91, Stufflebeam 84]. Although currently the attention continues to be focused on the practice, the knowledge provided by these theories helps to identify the components that the evaluators need to develop in order to execute an evaluation. Also, these theories allow us to classify the current evaluation methods. As an example, Table 1 shows the classification proposed by Worthen, Sanders, and Fitzpatrick in the social science program evaluation context [Worthen 97]. These types of evaluation methods can be applied to evaluate a program defined as "a complex group of people, organization, management, and resources that collectively make up a continuing endeavor to reach some particular educational, social, or commercial goal."

Besides this classification, the discipline of evaluation has also identified several evaluations that are performed in very different fields and disciplines. In Scriven's general vision of this study, he identified diverse categories of evaluations considering, for example, the main types of problems in the investigative disciplines (task-oriented classification of the evaluations, called the "Big Six," shown in Table 2) and the different types of evaluations performed in many specialty fields (type-oriented classification, shown in Table 3) [Scriven 00]. Some of the evaluations shown in Table 3 are well-established subfields of the "Big Six" and/or some overlapping with the task-oriented category.

Evaluation Approaches	General Purpose of Evaluation
objective-oriented evaluation	determining the extent to which goals are achieved
management-oriented evaluation	providing useful information to aid in making decisions
consumer-oriented evaluation	providing information about products to aid in making decisions about purchases or adoptions
expertise-oriented evaluation	providing professional judgments of quality
adversary-oriented evaluation	providing a balanced examination of all sides of controversial issues, highlighting both strengths and weaknesses
participant-oriented evaluation	understanding and portraying the complexities of a programmatic activity, responding to an audience's requirements for information

Table 1. Evaluation Methods Proposed by Worthen et al.

product evaluation		
performance evaluation/assessment		
personnel evaluation		
proposal evaluation		
program evaluation		
policy analysis		

Table 2. "Big Six," Task-Oriented Classification of Evaluations Methods

curriculum evaluation	literary, art, and music criticism	logic
technology assessment	psychological and medical assessment of patients	movie and restaurant reviewing
applied logic	the evaluation of the state of disciplines and academic areas	wine tasting
industrial quality control	appellate court jurisprudence	real state appraisal
the evaluation of research	perfume, sensory, and food evaluation	
diamond grading	investment portfolio evaluation	

Table 3. Type-Oriented Classification of Evaluation Methods

After analyzing the diverse components described for each type of evaluation, a set of elements is obtained that can be classed as basic, because they are common to any type of evaluation method (implicitly or explicitly identified, informally or formally described). The components could be denominated differently, although they refer to the same concept, according to the discipline in which the evaluation is executed and the type of method. These basic components are the foundations upon which the framework to elaborate an evaluation method was developed.

2.2 Framework to Develop an Evaluation Method: Evaluation Components

Figure 1 shows the evaluation components employing the generic nomenclature used in this document. Each component is described as follows:

- target: the object under evaluation
- criteria: the characteristics of the target that are to be evaluated
- yardstick or standard: the ideal target against which the real target is to be compared
- data-gathering techniques: the techniques needed to obtain data to analyze each criterion
- synthesis techniques: techniques used to judge each criterion and, in general, to judge the target, obtaining the results of the evaluation
- evaluation process: series of activities and tasks by means of which an evaluation is performed

As shown in Figure 2, all these components are closely interrelated. The evaluation can be customized by means of the target, because this is one of the parameters used to select the evaluation method. Once the target is known and delimited, its characteristics must be identified for evaluation (criteria). All the characteristics and their ideal values, which indicate what the target should be like under ideal conditions (or simply, under certain circumstances), make up what is known as the yardstick or standard. Data about the real target should be obtained using certain data-gathering techniques: a value (numerical, data, information set, etc.) will be gathered for and assigned to each criteria. Once all the data have been collected, they are organized in an appropriate structure and compared against the yardstick by applying synthesis techniques. This comparison will output the results of the evaluation. Finally, all of the above are linked by the evaluation process, which indicates when to define the scope and extent of the evaluation and when to develop, or adapt (if already available) and when necessary, the criteria, yardstick, and techniques. All of this is defined by a set of performance activities and tasks.
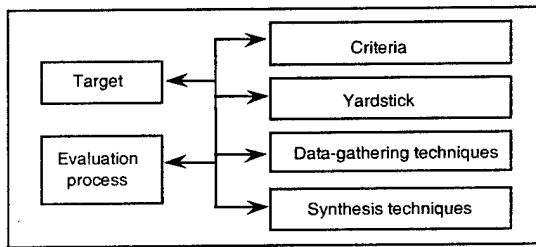
Figure 1.   Components of an
             Evaluation
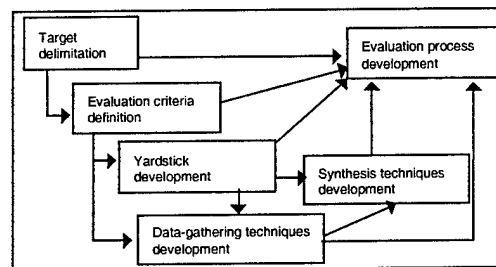
Figure 2.   Interrelations Between
             Evaluation Components

For each type of method considered, all the components must be defined explicitly for the evaluation to be conducted rigorously. This means that evaluators will know why a task must be performed at a particular time; this knowledge will help them to understand the evaluation method and components, and hence, eradicate or at least minimize interpretations, which always arise when evaluators do not understand the evaluation process.

The application of these components to a particular evaluation involves, first, selecting the best-suited type of evaluation method considering mainly the target under analysis. It is possible to use an eclectic approach based on diverse types of methods. The selection of a particular type of method determines some characteristics of the evaluation components, but does not mean, in all disciplines and fields, that the method is already developed or that there are specific guidelines that define how it should be developed. It merely allows certain types of techniques to be selected for use in designing each evaluation component and in specifying a set of characteristics for the method to be developed. For example, an objective-oriented evaluation is based on determining whether the target has met its goal or set of goals [Scriven 91].

The evaluation components shown in Figure 1 are the basis for developing the framework. We can develop a method that will be an instantiation of the application of this framework. However, if the framework is used by other evaluators, the final result (evaluation method) will not necessarily be exactly the same, because instantiation involves making a series of decisions, which may differ depending on the opinions and the environment of the evaluation method developer (e.g., considering other criteria or techniques). The following subsections outline the guidelines for developing each component, the techniques that can be used in each component, and an example focused on a particular evaluation (evaluating PC computers for buying, in both an informal and formal approach). We call the exposed guidelines a framework because these components are not particularized to a specific discipline or field. Also, we say that the result is an evaluation method because we are looking for a way to define explicitly and rigorously the evaluation components before executing an evaluation, and allowing other evaluators to perform another evaluations using the same components (criteria, yardstick, etc.) described in the method. However, we can find all these components in an evaluation performed in a non-formal way. That is why we clarify the meaning of each component by including in this document examples of both informal and formal evaluations.

2.2.1 Target

Target delimitation is the first essential step in any evaluation. In order to identify the criteria (discussed in Section 2.2.2), it is necessary to study in detail the object under evaluation and to delimit, in a general way, the factors to be considered. There are few techniques available for performing this step. Indeed, it is the experts in the field who very often indicate which factors are to be considered. However, evaluators can apply the functional analysis technique, described as the general description of the target's function. Depending on what the target is, evaluators can complete this analysis with the description of the context, stage of development, expected effects, and any other information that can help the evaluators under-

stand what the target is and delimit explicitly what will be analyzed in this evaluation. Sometimes the functional analysis technique is also used to identify the criteria of the evaluation.

The proposed example (personal evaluation for the purchase of a computer) is the typical evaluation performed when we want to buy a computer for use at home. In this case, the person who needs the computer is accustomed to focusing on the analysis of PC computers with the basic I/O components (monitor, keyboard, mouse, speakers, modem, and printer). In this specific case, a functional analysis of the target is not necessary because the target is not complex, and a mere general identification and description is enough to delimit what will be evaluated. In order to simplify the example, we will suppose that the computers have a fixed configuration (certain processor, certain memory, I/O devices, etc.) that we could not change. In a real evaluation, we would have to consider the possibility of analyzing each component (memory, processor, printer, etc.) separately to assemble all of these elements and obtain a computer more adapted to the needs of the person for whom the PC is being purchased.

In a more formal way, perhaps we need to develop a method of evaluation with the purpose of knowing which computer is the most appropriate for a person. In this case, as developers of a method, we have to focus on the description of the target, the explicit statement of its main characteristics (to be evaluated), and any other information that can be included to describe the context and scope of the method. This will allow the person (as the executor of an evaluation) to know if this method will fit with his/her needs. For example, if the person wants to consider the possibility of analyzing each computer component (memory, keyboard, etc.) independently, and the method does not include this option, the person will have to select another method or adapt the one above to consider this alternative.

2.2.2 Evaluation Criteria

Criteria definition is the second essential and critical step for developing a method of evaluation. Having ascertained and delimited the target, it is necessary to identify what characteristics of the target are of interest for evaluation purposes. These characteristics are referred to as evaluation criteria. In some cases, the evaluators have to use certain obligatory standards that contain implicitly the criteria to be applied in the evaluation (legal standards, professional standards, scientific standards, etc.). In other cases, diverse techniques for criteria elicitation can be used. The selection of the technique(s) to apply will depend on the target to evaluate and, in some cases, on the evaluation method selected. In general, evaluators can use the following techniques for criteria elicitation [Scriven 00]:

- functional analysis of the target: defined as the detailed description of the target's function

- needs assessment: refers to any study of the needs, wants, market preferences, values, standards, or ideals that might be relevant to the target

- complex logical analysis: when the definition needs more unpacking in order to figure out its implications. This is more often the case when the criterion is significance related. The analysis is a complex inferential process starting from data and definitions.

All of these techniques can be complemented using the basic set of key questions: what, why, when, how, where, and who. The purpose is to complete the analysis of the target and to assure that the target has been studied in detail. Usually, two types of responses can be gained from these questions: general criteria (characteristics that cannot be assigned a value directly and require further decomposition to which the set of questions will be applied successively until specific criteria are obtained) and specific criteria (characteristics that can be assigned a value directly using a particular data-gathering technique). Due to the fact that these responses are interrelated, it is possible to draw a diagrammatic tree that contains all the general and specific criteria that are to be evaluated. This criteria tree is the basis for developing the evaluation yardstick (Section 2.2.3) and for selecting the data-gathering techniques (Section 2.2.4). As a complement to this tree, the description of each criterion should be added, including its specific meaning and if it is an isolated characteristic (stand-alone criterion) or if it is related to other criteria (compensatory criteria). In this way, the tradeoffs among criteria could be expressed, if they exist in the target being considered. Furthermore, the criteria tree and these definitions aid in the accurate understanding of the yardstick, because the evaluator will know exactly what characteristics are to be analyzed.

Regarding the example, usually the person for whom the PC is being purchased does not write the criteria to be analyzed and is not conscious of the application of criteria elicitation techniques. Normally, a set of characteristics derived from personal reflections is selected. Since in the example the same person is the designer of the evaluation, the evaluator, the user of the evaluation results, and the purchaser of the PC, the analysis of the needs shown below can be considered correct:

> *"I need the computer to work with my photographs (I have a digital camera) and graphics of my work. Therefore, the computer must have appropriate processor speed, hard-disk capacity and RAM for this type of applications; furthermore, I need a monitor of high quality, with an appropriate graphic card and a color printer. Finally, as I am left-handed, I want to have a keyboard and a mouse adapted for left-handed persons."*

In this reflection, we can easily identify the criteria that this person will consider in the evaluation: processor speed, hard-disk capacity, RAM, monitor, graphic card, printer, keyboard, and mouse. However, if we are method developers, we have to write down all the criteria and describe them explicitly, as a basis to develop the next evaluation components. The above informal description and the target delimitation are the main inputs to decompose the general characteristics into specific criteria. For example, a printer is a general criterion that can be broken down into other specific criteria. We do not want to analyze only whether the computer has a printer; on the contrary, we want to know about its speed, dimensions, and type, for example. With this set of characteristics, we can develop the criteria tree, a table describing each criterion, and a list of the any tradeoffs. Figure 3 shows an example of a criteria tree and lists the meaning of each criterion on it. At this point of the evaluation method development, we do not have to analyze which are the "appropriate" or "ideal" val-

ues for speed, capacity, and so on. Now we are interested only in the identification and description of the criteria that will be analyzed.
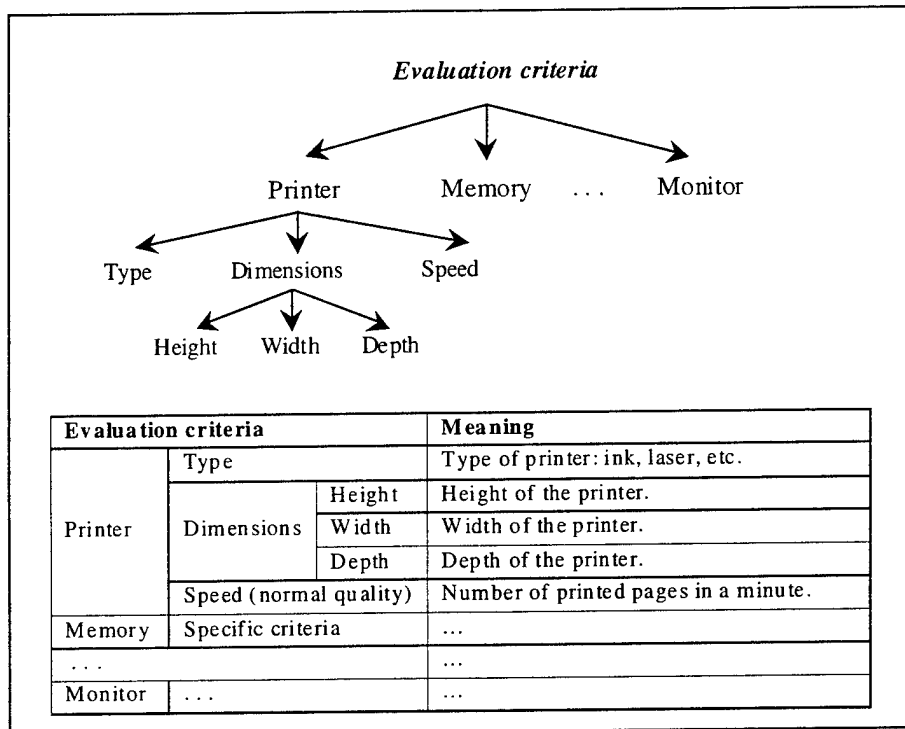
Figure 3. Example of Criteria Tree and Criteria Description

2.2.3 Yardstick

The description of the target and the criteria tree are the basis for developing the yardstick. Depending on the discipline and target considered in the evaluation, we could find different types of yardsticks (prescriptive standards, descriptive narration, etc.). All yardsticks must contain the specifications, requirements, descriptions, or values for each criterion considered. So, if evaluators have to develop the yardstick, they can use the following bases:

- The yardstick used in the evaluation should be developed from the criteria tree obtained in the preceding step. The general structure of the yardstick should be inferred from the criteria tree.

- The yardstick must contain the specifications of all defined criteria.

- For each criterion, whenever possible, the yardstick must define the specifications structured as pairs [criterion, datum/information].

- Whenever applicable, the yardstick must contain threshold values to indicate the minimum value for each criterion to be reached for a positive evaluation. For example, in an academic context, we can evaluate the performance of the students on a subject using five grades (A through F) and defining the threshold value D as the minimum value that students must have to pass the subject. This task is closely related to synthesis tech-

niques. Due to this, these threshold values could be defined when the evaluator has se-
lected the synthesis techniques to be applied in the evaluation.

In the PC-purchase example, once the criteria set has been elicited, we have to assign a value
(quantitative or qualitative) to each criterion. Table 4 shows sample values that could be as-
signed to the criteria elicited. Usually in an informal evaluation, these values are not written
down in any document, although the person knows the "ideal" values of the "perfect" com-
puter. When developing an evaluation method, we have to describe all of the yardstick and, if
needed, include information to explain the values assigned (i.e., if threshold values are in-
cluded, descriptions and/or examples to understand the meaning of the threshold) and, if there
are tradeoffs between criteria, data and/or examples to understand these relationships.

Evaluation Criteria			Yardstick
Printer	Type		Ink (black and color)
	Dimensions	Height	9-12 inches with paper tray up
			14-22 inches with paper tray extended
		Width	15-25 inches
		Depth	7-15 inches
	Speed (normal quality)		3-7 ppm
Memory	Specific criteria		...
. . .			...
Monitor	. . .		...

Table 4. Example of a Yardstick

2.2.4 Data-Gathering Techniques

Apart from building the yardstick, the potentially applicable data-gathering techniques need
to be identified, and one or more need to be assigned to each evaluation criterion. The objec-
tive of applying these techniques is to obtain the information needed to judge the target with
the next component (synthesis techniques). The main data-gathering techniques used in most
evaluations in the software engineering field can be classed in three groups, as shown in
Table 5. Many of these techniques are also used in other disciplines. For example to evaluate
social programs, the most common techniques are: survey, interview, test, observation, group
techniques, case study, photograph, videotape, slides, document review and analysis, portfo-
lio review, testimonials, expert or peer review, simulated problem or situation, journal (or log
or diary), and unobtrusive measures [Taylor 96].

Type	Description
measurement	involves the use of the appropriate measurement instruments or mechanisms
assignation	for example, questionnaires, interviews (individual or groups), documentation inspection, and simple tests not involving metric applications
opinion	techniques for getting subjective criteria data, such as, by observation

Table 5. Types of Data-Gathering Techniques

The selection of data-gathering techniques will depend on the preceding components (target, criteria, and yardstick). One or more data-gathering techniques must be assigned to each criterion by analyzing the meaning of the criterion and the type of value (numerical, data, etc.) specified for it in the yardstick. Once identified and assigned, each technique must be developed, outputting questionnaires, standard interviews, lists of documents for inspection, observation forms, metrics, and so on. It is also recommended to attach examples of the practical application for each technique.

In the PC-purchase example, after identifying the criteria and developing the yardstick, the person thinks about how to obtain information about all the computers to be analyzed. Usually, the gathering of data is carried out through the analysis of informative pamphlets or booklets, asking sellers and, if it is possible to perform in all the cases (to avoid bias), carrying out tests: for example, prove the real speed of the microprocessor in a given situation and benchmark. If thinking about an evaluation method, the evaluators will assign, for each criterion, a data-gathering technique to obtain data taking into account: the meaning of each criterion; the type of value specified in the yardstick; and the possible techniques that the individual (in the evaluator role) can apply. For example, to analyze the printer speed the person can check the printer documentation (provided by the supplier), ask the seller about the printer (or another person that works with the same printer model), or perform a printer test to control the number of pages printed per minute. Table 6 shows a sample assignment of data-gathering techniques. To develop an evaluation method rigorously, the evaluators would select a technique for each criterion and develop all the selected techniques. The data/information gathered after applying these techniques can be written down in a table just like the yardstick table (see Table 4) for each analyzed computer. When developing the evaluation method, these techniques are identified, assigned, and developed, but not applied. As method developers, we are interested only in their explicit description.

Evaluation Criteria			Data-Gathering Techniques
Printer	Type		document review, observation
	Dimensions	Height	document review, test, interview
		Width	document review, test, interview
		Depth	document review, test, interview
	Speed (normal quality)		document review, test, interview
Memory	Specific criteria		...
...			...

Table 6. Sample Assignment of Data-Gathering Techniques to Criteria

2.2.5 Synthesis Techniques

Synthesis techniques are used to synthesize all the data and information obtained after applying the data-gathering techniques and for comparison against the yardstick (i.e., to judge the target and obtain the results of the evaluation). Usually, two types of synthesis techniques can be applied:

- single value: a single datum (numerical or otherwise) is obtained as a result of the evaluation. This group includes combination methods. When these techniques are applied, a meaningful value scale is required for the datum obtained.

- multiple values: These techniques, for example, statistical techniques, criteria grouping, and datum-by-datum comparison with the yardstick, output more detailed information than single-value techniques.

As was the case with the data-gathering techniques, the selection of the synthesis techniques will depend on the preceding components. Single-value techniques are required, usually, for comparative evaluations. For example, by obtaining a final value for each evaluated computer we will decide easily which of the computers is the winner. Multiple-value techniques are applied in most cases. For example, criteria grouping and datum-by-datum comparison with the yardstick are the most frequently used techniques in the improvement-oriented evaluation methods. But independent of the selected synthesis techniques, describing how to synthesize the information using examples is recommended to develop each technique completely. This is important when there are compensatory criteria (tradeoffs between criteria) and can help determine how to obtain the final result. As Scriven said:

> "The question is, What exactly is the rule that explains how the tradeoffs are made and the overall result judged? If there is no rule, which means it's a judgment call, then the results will typically vary depending on who happens to be doing the judging on a particular day" [Scriven 00].

In the PC-purchase example and other informal evaluations, after data gathering, the person has all the information needed to judge which computer best satisfies his or her identified needs. In this case, the synthesis of data depends on the judgement of the individual: there are

no rules or algorithms to obtain the final result. When developing an evaluation method, these techniques would be defined explicitly. For example, to evaluate which is the best computer, the evaluators could use the numerical weight and sum model for each computer. In this model, the criteria are weighted for their relative importance (e.g., on a 1-3, 1-5, or 1-10 scale), and then points are awarded for the merit of each computer candidate's performance on each of these valued criteria (e.g., on a 1-5 or 1-100 scale). The products of the weights and the performance scores are calculated and totaled for each candidate: the best candidate being the one with the highest total. As developers of an evaluation method, we are interested in the description and elaboration of the synthesis techniques but not their application or execution (i.e., we will describe the algorithm to apply and all the exceptions and rules to be considered when deriving the evaluation results).

2.2.6 Evaluation Process

The evaluation process is a series of specific activities and tasks that are to be executed to perform an evaluation. All the previous components are necessary to describe and design an evaluation method, but it is the evaluation process that describes the list of activities to perform and when to use the previous elements in practice. The framework describes three main subprocesses (shown in Figure 4) that match the three major points through which an evaluation passes: prepare the evaluation, get data about the target, and make a decision on which the final report of the evaluation will be based. These subprocesses are a generalization of the evaluation processes proposed by the basic methods of evaluation, which are usually tailored to particular fields:

- planning or preparation: activities involving making contact with the target to be evaluated, delimiting the evaluation, and planning and managing its execution. This phase ends when all the evaluation components have been developed (or, if necessary, adapted to the target in question) and the team of evaluators is ready to make a visit to or interact with the target.

- examination: application of the data-gathering techniques and obtaining the data required to judge the target. This phase ends when all the information has been obtained for all the criteria considered in the evaluation.

- decision making: application of the synthesis techniques and development of the final report. Also, this activity includes the task of completing the documentation of the evaluation, whose goal is double: to refine the evaluation process (and therefore the method of evaluation) and to maintain the documentation of this evaluation to compare it with future evaluations of the same or similar target.

	PLANNING
	• Establish the evaluation goals.
Planning	• Design the evaluation.
	• Analyze the target.
	• Plan the evaluation.
	EXAMINATION
Examination	• Apply the data-gathering techniques and obtain the data needed.
	• Check the data gathered for completeness.
	DECISION MAKING
	• Apply the synthesis techniques.
Decision Making	• Prepare the final report.
	• Present and submit the final report.
	• Complete the evaluation documentation.

Figure 4. Main Subprocesses and Activities of the Evaluation Process

The specific activities and tasks to be performed in a particular evaluation will depend also on the precedent components and their particular development. For example, if the data-gathering techniques include the possible use of metrics (but only if the mechanisms to obtain them are developed and applied), the evaluation process would include the activities needed to determine if these mechanisms are implemented and if it is possible to use this technique to obtain the data. If metrics will not be used, these activities will be omitted.

As this general evaluation process will be the point of reference for comparison with the ATAM process, each subprocess needs to be described in more detail. This description is given in Table 7, in which DO is the development organization, and EO is the evaluation organization.

A. Planning Subprocess	
A.1. Establish the evaluation goals.	This activity kicks off when contact is made with the DO and is concluded when either the evaluation contract is drafted and signed or the evaluation is cancelled.
A.1.1. Get to know and analyze the DO target.	The evaluation will start when the DO requests an evaluation. The EO will select a representative or a small group of evaluators whose job is to run a preliminary analysis of the target to determine whether the DO can provide the information required to run the evaluation. If the DO does not have the data required, the evaluation should be cancelled.
A.1.2. Negotiate with the representatives of the DO.	During these contacts with the DO, the representative of the EO (or small group) will describe the method of evaluation to be applied and delimit the target together with DO representatives.
A.1.3. Define the goals of the evaluation.	Taking the results of the above activities as a basis, the EO representative (or small group) will define the goals of the evaluation, use of evaluation results, and a very rough schedule, as well as budget and time estimates. The main output of negotiating these goals with DO representatives will be either the evaluation contract (supposing that both organizations agree) or the decision to cancel the evaluation, if no satisfactory agreement is reached. If agreement is reached, a schedule will be developed for the other activities of this subprocess.
A.2. Design the evaluation.	The evaluation components to be used in the evaluation in question must be detailed and designed. Depending on the discipline and field in question, this may mean developing all the evaluation components (if none exist) or adapting the components defined in the evaluation method selected (where any type of modification of any component depending on the target for analysis will be considered as "adaptations"). Should no adaptation at all be required, this activity can be skipped. As an output, all the created/adapted/selected components will be included in the "Evaluation Design" document. In some cases, if it is necessary to get more information about the target in order to design the components, the evaluation must be designed simultaneously with the following activity.
A.3. Analyze the target.	The results of all these tasks will be used to completed the "Evaluation Design" document, which was begun in the preceding activity.
A.3.1. Request and analyze the general DO documentation.	This information is required to determine the number of evaluators required and to develop the data-gathering and synthesis techniques.
A.3.2. Set up the full evaluation team.	When necessary, train the evaluation team.
A.3.3. Identify the professionals involved.	Identify/select the DO professionals involved in the evaluation. Assign roles to each person. Identify the evaluation report addresses.
A.3.4. Develop the data-gathering and synthesis techniques.	If the selected evaluation method does not provide these techniques, the evaluation team has to develop them. In other cases, the team must determine if it is necessary to adapt these techniques based on the target, criteria, and yardstick.
A.3.5. Develop/adapt the infrastructure.	Develop (or adapt, if already provided by the method) the infrastructure to be used to gather the information (computer media, tables, etc.).

Table 7. Description of the Evaluation Process Activities and Tasks

A. Planning Subprocess (cont'd.)	
A.4. Plan the evaluation.	Considering the above tasks and evaluation components, all the activities and resources required will be planned in detail. A manager from the DO must be involved to plan tasks in which DO professionals are involved and to negotiate and accept the plan and final costs of the evaluation. After acceptance, the EO representative will present the method to the DO professionals, paying special attention to the tasks that are to be performed during the visit to the organization. Some of the tasks of the "plan evaluation" activity can be run simultaneously with the preceding activity and refined as more information is gathered.
B. Examination Subprocess	
B.1. Apply the data-gathering techniques and obtain the data needed.	The data-gathering techniques selected and developed in the planning subprocess will be applied in this activity. The type and number of techniques for each application depends on the particular evaluation being run.
B.2. Check the data gathered for completeness.	This activity can be run simultaneously with the activity above and may even call for modification of certain data-gathering techniques, if any of the following is true: • Any datum necessary for running the evaluation is missing. • There are inconsistencies or contradictions in the data supplied by some stakeholders. • There are omissions, meaning that the majority of stakeholders has not responded to some questions. All the information gathered will be compiled and attached to the final evaluation documentation.
C. Decision-Making Subprocess	
C.1. Apply the synthesis techniques.	After gathering the information required for each criterion, the synthesis techniques will be applied to judge the target. This activity is concluded when all the selected criteria have been judged, and the results of the evaluation have been obtained.
C.2. Prepare the final report.	Taking the results obtained, a final report is prepared. This report will be delivered to the evaluation recipients (identified in the planning subprocess). Some evaluations call for a range of reports to be developed depending on the responsibility and job of each recipient. This activity ends when all the required evaluation reports have been developed.
C.3. Present and submit the final report.	Some methods call for the evaluation results to be presented to the stakeholders at a meeting. For this purpose, the evaluation team will have to prepare the presentation on the basis of the report output. According to Stufflebeam, the presentation of the report would be the main duty of a specific evaluation team role: a specialist in communication [Stufflebeam 84]. The evaluation team will have to apply the strategy selected in the evaluation design to circulate the report.

Table 7. Description of the Evaluation Process Activities and Tasks (cont'd.)

C. Decision-Making Subprocess (cont'd.)	
C.4. Complete the evaluation documentation.	An evaluation is usually concluded when the report has been presented and distributed to its addressees. However, the evaluation team must compile all the information/documentation used and/or generated to do the following:
	• Assure that the evaluation documentation is available for comparison with the results obtained in future evaluations.
	• Refine the evaluation method. The evaluation manager will be in charge of analyzing the processes and techniques applied and comparing them with the method definition. Improvements to the evaluation method can be made, if necessary, after analyzing the schedule deviations, differences in the order in which activities were performed, modification of any technique, and so on.
	An evaluation is really concluded when all the information/documentation has been compiled. Evaluation method refinement is an activity that should be performed by the method developers, who are not necessarily the same people as the method executors.

Table 7. Description of the Evaluation Process Activities and Tasks (cont'd.)

In the PC-purchase example, in the informal context, when all the previous components have been considered (although they are not written down in any document) the computers that will be analyzed must be selected. Next, data will be gathered (application of the data-gathering techniques), and each computer will be judged (application of the synthesis techniques) to finally obtain the best candidate. When developing an evaluation method in which diverse evaluators are usually involved, it is necessary to describe the activities to be performed in a more formal way. For example, Table 8 shows an evaluation process that should be applied in this case.

A. Planning Subprocess	
	Establish the evaluation goals. Between the stakeholder(s) and the evaluation team, define the goals and boundaries of the evaluation and the use of the evaluation results. Negotiate and decide whether the evaluation will be performed.
	Design the evaluation. Although the evaluation method is developed, it may be necessary to adapt the techniques to the real target that will be analyzed: for example, selecting a certain group of data-gathering techniques (among all the techniques proposed by the method) because it is not possible to carry out tests in all the computers to be analyzed. In other evaluations when it is not possible to analyze all the targets, the evaluators have to select a representative sample; as a result, only a subset of targets will be analyzed. In this case, the evaluation method has to describe the sample technique explicitly.
	Analyze the target. In this case, there is no need to obtain information from the DO. However, it is necessary to set up the evaluation team, train them (if necessary), assign roles, develop the data-gathering and synthesis techniques, and develop the infrastructure to gather the information (if the method does not provide it).
	Plan the evaluation, in order to manage all the necessary resources and activities to be performed. Assign roles to each individual (not evaluators) involved in the evaluation (for example, clients, end users, etc.).
B. Examination Subprocess	
	Apply the data-gathering techniques and obtain the necessary data to be used by the synthesis techniques.
	Check the data gathered for completeness. If any data are missing, some of the techniques applied will be modified as appropriate in order to account for all the information required to evaluate the target. With the purpose of improving the evaluation process itself, the following should be recorded: all the incidences detected, any activities performed in a different way or in a different order, duplicated tasks, and so on.
C. Decision-Making Subprocess	
	Apply the synthesis techniques. Synthesize the gathered data to obtain the final results of the evaluation.
	If required, prepare the final report.
	If required, present the final report at a meeting and circulate the report to its addresses (via email, technical report, general publication, bulletin board, etc.).
	Collect all the information and documentation generated in the evaluation in order to refine the evaluation process and maintain the documentation for future comparisons with the results of other evaluations.

Table 8. Evaluation Process for Evaluating Computers

Therefore, the evaluation process specifies when to apply the previous evaluation components. However to develop or to perform an evaluation, it is necessary to take into account all of the evaluation components: target, criteria, yardstick, data-gathering techniques, synthesis techniques, and evaluation process. All of these concepts will be the starting point for analyzing the ATAM in the next section.

3 Analysis of the ATAM

The ATAM evaluation method was developed by the Software Engineering Institute (SEI) to evaluate specific architecture quality attributes and engineering tradeoffs to be made among possibly conflicting quality goals [Jones 99]. As shown in Figure 5, the key input to the ATAM is an architecture and the main outputs include the following:

- risks, or architectural alternatives that might create future problems in some quality attribute

- non-risks, or good decisions relying on implicit assumptions

- sensitivity points, or alternatives of which a slight change makes a significant difference in some quality attribute

- tradeoff points, or decisions affecting more than one quality attribute [SEI 00]
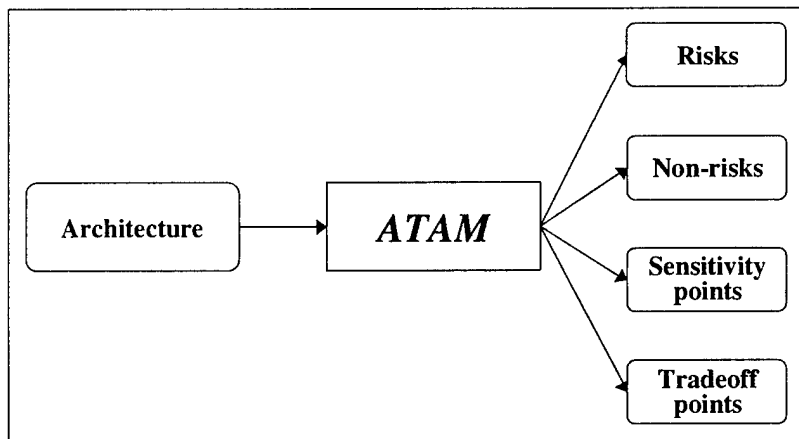
Figure 5. Main Input and Outputs of the ATAM

The ATAM can be used in different situations: to create pre-concrete architecture definitions at the discovery phase; to analyze architecture decisions, with little or no code; to analyze alternative candidate architectures; and to evaluate existing systems [SEI 00, slide 45]. Nevertheless, it is necessary to highlight that the ATAM is intended to "analyze an architecture with respect to its quality attributes, not its functional correctness" [Jones 99]. Also, accord-

ing to the *ATAM Reference Guide*[1], it is not a code evaluation and does not include any actual system testing. The ATAM is applied by an external evaluation team not related to the DO.

In this section, the ATAM will be analyzed to identify the basic evaluation components, described in Section 2 on page 3. This analysis, which will help describe the evaluation method, is based on three basic references: the *ATAM Reference Guide*, the SEI ATAM presentation [SEI 00], and *Attribute-Based Architectural Styles* [Klein 99]. Other references will be used if the basic documentation has not included the definitions or the development of certain evaluation components. One of these references will be the recent ATAM report [Kazman 00], developed almost simultaneously with this analysis. The structure applied to present the results of this analysis, for each evaluation component, is the following:

- For the target, the main issues of the functional analysis of an architecture evaluation will be used to describe the target considered in an ATAM evaluation.

- With regard to the evaluation criteria and yardstick, the identification of the evaluation components and their elicitation (during the development of the evaluation method and the execution of an ATAM evaluation) will be addressed.

- The development of the data-gathering and synthesis techniques will be analyzed.

- The comparison between the ATAM evaluation process and the activities described in the framework (Table 7) will be included.

- Suggestions are included to clarify each component.

Figure 6 shows the phases and steps of an ATAM evaluation (as documented in the *ATAM Reference Guide*) and shows the pair of phase and step numbers that identify each one.

---

[1]   The *ATAM Reference Guide,* written by the staff of the Software Engineering Institute's Attribute Tradeoff Analysis Initiative, has not yet been published.
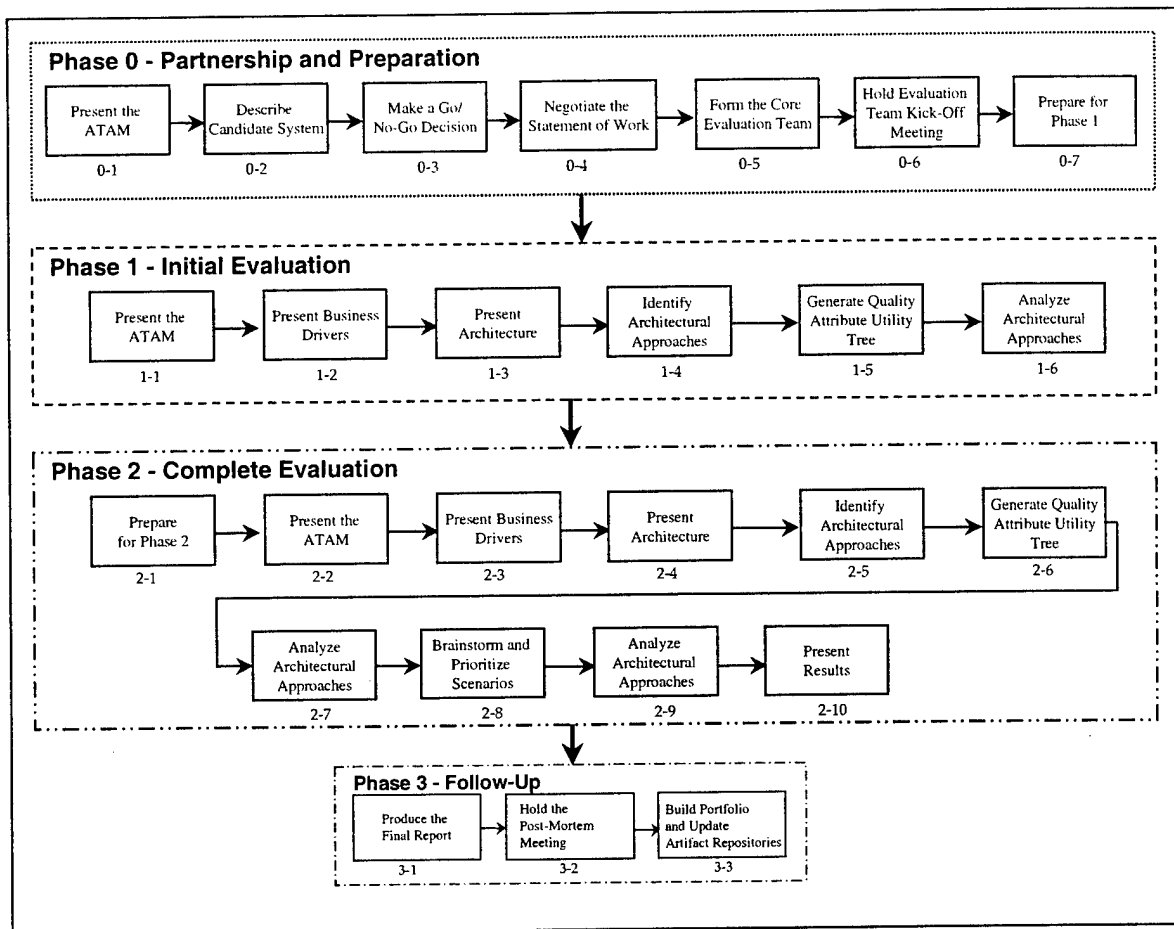
Figure 6. ATAM Phases and Steps

3.1 Analyzing the Target

The outline shown in Figure 7 was developed to analyze the target of the ATAM. This outline contains a set of important elements that should be considered to identify and delimit the target. These elements would be the result of a functional analysis and have been selected because they can be used to get at least the essential information for understanding what is to be evaluated by the ATAM. This outline will be used in the next sections to identify and analyze the general definition and delimitation of the ATAM target and the main factors for evaluation.

1. What is to be evaluated ...
2. The target can be defined as ...
3. Classed in software engineering (SE) under ... and, therefore, a concept applicable to the paradigm(s) ...
4. The target needs to be evaluated because ...
5. The main target factors for consideration are ... and these factors need to be evaluated because ...

Figure 7. Main Issues in the Functional Analysis

3.1.1 Identification of the Target

The target of the ATAM is an architecture. However, the basic references fail to distinguish explicitly between software architecture, system architecture, and software system architecture and do not indicate whether the words *system* and *software* are considered synonymous. By way of an example, consider the following paragraphs taken from diverse references.

ATAM Reference Guide	"Early evaluation of the *architecture of a system* or a product line of systems is a low cost, risk reduction method... ATAM consists of a *system* or product line architecture and the perspectives of stakeholders involved with that system or product line..."
[SEI 00, slide 3]	"*Software* architecture is the structure or structures of the system..."
[SEI 00, slide 20]	"We do architecture analysis to ... (1) compare high level designs for a *system* and document those comparisons... A *software* architecture is the earliest life-cycle artifact that embodies significant design decisions: choices and tradeoffs."
[Kazman 00, p. 2]	"The ATAM is meant to be a risk identification method, a means of detecting areas of potential risk within the architecture of a complex software-intensive system."

Bass et al. state that "System architectures are not the main focus of this book, but at times we need to show the allocation of software to hardware to understand a software architectural design decision" [Bass 98, p. xvii]. However, the same book in which that quotation appears contains a host of references to system architecture, references that are sometimes used as synonyms of software architecture, whereas at other times, the terms are confused, and it is very difficult to tell from the context whether the term refers to the software architecture or to the system architecture (especially when the authors have described "a visual notation for presenting software and systems architectures" [Bass 98, p. xvi]). From these references and the definition of the target given by Bass et al. and in the SEI ATAM presentation [SEI 00, slide 3], we will assume that the ATAM target is the software architecture. However, given the apparently indistinct use of the terms *system* and *software architecture* in other references (including the basic documentation), the target can be confused with the architectural design with which system engineering, not software engineering, is concerned. One could possibly reach this conclusion, taking the following into account:

- current research about the integration of system engineering and software engineering [Boehm 00, CMMI 99]

- The fact that the relationship between the software architecture discipline and system engineering or software engineering is not described explicitly, although Bass et al. specify that "we will focus on architecture strictly from a software engineering point of view" [Bass 98, p. 21].

- "The system/software architect will be asked to make a presentation that explains the architecture for the system" (according to the *ATAM Reference Guide*). However, according to Kazman et al., "the system architect presents the architecture to the evaluation team" [Kazman 00].

Therefore to delimit the target, it is prudent to identify explicitly the target of an ATAM evaluation. In the next subsection, the target (software system architecture) will be analyzed within software engineering, taking into account its relationship with the diverse current software development paradigms.

3.1.2 Definition of the Target

According to the SEI ATAM presentation, "software architecture is the structure or structures of the system, which comprise software components, the externally visible properties of these components, and the relationships among them," and "The exact structures to consider and the ways to represent them vary" [SEI 00, slide 3]. For a more detailed analysis of the target, we have to consult Bass et al. who include a dissertation on this recent discipline and the ensuing problems arising out of disagreement on what software architecture is and the structural issues it should contain. A list of the most common and useful software structures is also provided, which "together describe the building's architecture," as there is no *one* structure that is the software architecture [Bass 98, p. 36].

According to the *ATAM Reference Guide*, the following views—functional, module/layer/subsystem, process/thread, and hardware—will have to be described in the presentation given by the architect in order to perform an ATAM evaluation. However, this reference fails to further describe the relationship of these views with quality attributes. We have to consult Bass et al. to find this association, although the number of views is not exactly the same [Bass 98, p. 38]. The views described by Bachmann et al. (logical, concurrency and deployment), which will be the views obtained if we apply the architecture-based design method, also differ [Bachmann 00].

However, although these views are identified, the method does not describe explicitly the documentation (not information gathered through presentations) required to run the evaluation. Nevertheless, bearing in mind research team experience, it would be a good idea to include an example of a software architecture specifying the different views required by the method. This would provide a range of benefits: ease method training; ease stakeholder understanding of what is to be evaluated and, consequently, when it would be ready to request the evaluation; and allow for a more rigorous and formal definition of the evaluation method by specifying the target in detail.

3.1.3 The Target in Software Engineering

The relationship of the target with the different software production processes now defined in the diverse SE paradigms or life cycles is not specified explicitly in the basic references. From the following statements we can deduce that the software architecture is a product of the software system design phases: "we do architecture analysis to focus design activity where it is needed most" [SEI 00, slide 20]; and "an architecture evaluation can be done at a variety of stages during the design process" (according to the *ATAM Reference Guide*).

With respect to the development paradigms, no relationship of the target with any software production paradigm (structured, object-oriented, etc.) is described in any of the basic references. At first glance, given the nomenclature used in the ATAM (use cases, scenarios, etc.), this method of evaluation can be associated with object-oriented (OO) and with OO-software systems. However, Klein et al. state that this attribute-based architectural style (ABAS) "can also arise in object-oriented systems" [Klein 99, p. 10]. From this, we can deduce that ABASs are applicable to a range of paradigms and, therefore, the ATAM is also applicable to different paradigms. Nevertheless, if these relationships are stated in the method description, readers can easily know whether the method is applicable to all paradigms.

3.1.4 Need to Evaluate the Software Architecture

The justification of software architecture evaluation is addressed by the SEI ATAM Presentation and Bass et al. [SEI 00, Bass 98]. The above justification is included when generally describing the goal of the evaluation. For example, the importance of the software architecture may be justified for the following reasons [SEI 00, slide 5]:

1. "Software architecture represents earliest design decisions. These decisions: are hardest to change in the future; are most critical to get right; and are a communication vehicle among stakeholders."

2. "Software architecture is the first design artifact addressing four, at least, relevant quality attributes: performance, reliability, modifiability and security."

3. "Software architecture is the key to systematic reuse. This model is transferable and a reusable abstraction."

3.1.5 Main Factors for Analysis in the ATAM and Their Justification

Any of the basic references identify explicitly the main factors for analysis in the ATAM: quality attributes, like performance, modifiability, and availability [SEI 00]. These factors are general characteristics that cannot be measured or analyzed directly. Therefore, a structure which allows us to understand and analyze each attribute has been created. Based on this structure (called characterization of a quality attribute), the specific evaluation criteria to be considered in a particular ATAM evaluation will be elicited. However, it is important to stress that these quality attributes are interrelated and "their satisfaction can never be achieved in isolation" [Bass 98, p. 78]. The number of quality attributes for consideration is not identified

specifically in the basic references. The final set of quality attributes will depend on the stakeholders. With regard to justifying the selection of quality attributes, we have to consult Barbacci et al. who describe the different schools/traditions concerning the properties of critical systems (performance, dependability, security, and safety) and the best methods to develop them [Barbacci 99, Barbacci 95].

3.2 Evaluation Criteria

3.2.1 Identification of Evaluation Criteria and Their Structure

The "target" component provides the initial set of factors or general criteria: quality attribute requirements, such as performance and modifiability. As these criteria cannot be analyzed directly, the method's developers have created a structure to "characterize" each quality attribute. The quality attribute characterization is the key ATAM concept that can be associated with the evaluation criteria. There is another source of criteria in phase 0: during the negotiation between the DO and the EO, the DO should provide an initial description of the architecture and the quality attribute requirements that the DO thinks are more important (according to the *ATAM Reference Guide*). Nevertheless, this potential source of criteria will not be considered in this analysis due to the very general description of the activities related with these criteria; for example, the interrelationship between the criteria obtained after applying techniques like utility tree and the criteria identified in phase 0; and the activities which the evaluation team should carry out to perform a preliminary analysis of the architecture to be evaluated.

Each quality attribute characterization is composed of three elements: stimuli, architectural decisions, and responses [Kazman 00, p. 9]. Figure 8 shows these three elements as a graph (inverted tree) and gives a definition of each element. However, not all of the elements of Figure 8 can be catalogued as evaluation criteria. In the ATAM evaluation, the architectural decisions and responses are analyzed to determine risks, non-risks, sensitivity points, and tradeoff points. The stimuli only indicate the input variables that activate a response of the architecture related with certain architectural decisions. So, the stimuli are not criteria; they are elements required to identify a given variable (and its value), which will be the trigger for analyzing the architectural decisions and responses in a given setting.

Element	Description
Stimuli	Events that cause the architecture to respond or change.
Architectural Decision	Aspects of an architecture (components, connectors, and their properties) that have a direct impact on achieving attribute responses.
Responses	Measurable/observable quantities.
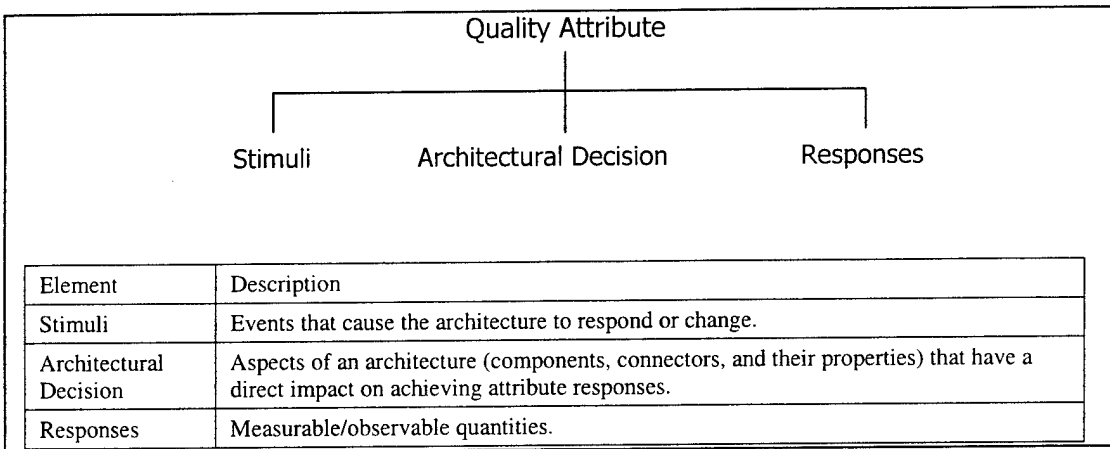
Figure 8. Quality Attribute Characterization Elements

Therefore, the triplet [stimuli/architectural decision/responses] contains the specific criteria of the evaluation (architectural decisions and responses) implicit in a structure by means of which a set of scenarios for analyzing the architecture can be derived easily. Taking into account that the ATAM identifies explicitly the characterizations of the quality attributes for analysis, the identification of the criteria can be said to be explicit.

3.2.2 Elicitation of Evaluation Criteria

The elicitation of specific evaluation criteria can be analyzed from two viewpoints: during the development of the evaluation method, and in the execution of an evaluation. With regard to the development of the method, the elicitation of the specific criteria is based on experts and their knowledge of the analysis of quality attributes in other communities.

With regard to method execution, as the target is not delimited specifically in the ATAM, nor is there a predetermined set of criteria for analysis or a yardstick serving as a point of comparison for the evaluation, some sort of technique needs to be applied to elicit the general criteria (quality attributes) and specific criteria (the architectural decisions and responses for each quality attribute) that will be considered in each particular evaluation. This elicitation is performed by means of two interrelated mechanisms: a utility tree and the identification of architectural approaches. Also, there is another potential source of criteria elicitation: brainstorming scenarios. It is classed as potential because its purpose is not criteria elicitation but scenario elicitation; once the scenarios are described, a quality attribute should be associated with each one. The set of quality attributes should be matched with the criteria identified in the utility tree. If the evaluators discover any criteria not considered before, the method does not describe how this affects the analysis performed, taking into account that brainstorming is one of the last steps of the evaluation. The identification of new criteria in the brainstorming step will not take place in all the evaluations, although it should be take into account to interrelate ATAM steps appropriately to give evaluators rules to apply in these cases.

Utility tree "is a top-down vehicle for characterizing the 'driving' attribute-specific requirements" [SEI 00, slide 33]. The specific criteria that the stakeholders wish to analyze in the architecture are detailed at the third level of the tree. A scenario that specifies an environment in which the criterion will be analyzed is associated with each specific criterion. All the scenarios will be prioritized (by the importance of each node to the success of the system and the degree of perceived risk posed by the achievement of this node) to get the set of general and specific criteria that, according to the stakeholders, will determine the success of the system. Nevertheless, besides these general guidelines, this technique is not described in more detail in the basic references, although Kazman et al. provide a more complete example [Kazman 00, p. 17].

The other mechanism for eliciting the specific criteria of an evaluation is related to the identification of architectural approaches. After finding out which quality attributes to analyze, the "evaluation team identifies approaches inherent in the evaluated architecture" (according to the *ATAM Reference Guide*). Based on the identified architectural approaches, the specific criteria to be evaluated are also obtained because each approach has an associated set of stimuli, architectural decisions, and responses that will be analyzed.

3.3 Yardstick

3.3.1 Identification of the Yardstick and Its Structure

The "evaluation criteria" component provides the set of criteria (generic and specific) for evaluation, described according to the triplet [stimulus/architectural decisions/responses]. This triplet is the basis for developing the ATAM yardstick: scenarios. The set of scenarios (that have been elicited, either implicitly or explicitly) in a particular evaluation is the reference point against which the architecture will be judged. Although there are diverse types of scenarios, the most important ones are the use cases, which represent the ways the stakeholders expect the system to be used; so, use cases represent a part of the requirements the evaluated architectural design must satisfy. As the ATAM is a scenario-based evaluation method, the structure of its yardstick is also based (albeit implicitly) on the general definition of a scenario: stimuli, environment, and responses. Therefore, the ATAM yardstick is a set of scenarios generated by three sources: utility tree, ABASs, and brainstorming. With the utility tree and brainstorming we can obtain scenarios explicitly whereas we can elicit scenarios not directly dependent on the stakeholders but on the architectural styles used in the architecture from the ABAS. On this basis, the different parts of the ATAM yardstick can be classed as:

- architectural design-dependent scenarios: ABASs

- stakeholder-dependent scenarios: generated by utility tree and brainstorming

  - The utility tree provides a top-down mechanism for eliciting a set of initial scenarios derived directly from quality attribute requirements to be analyzed. These scenarios, located in the leaves of the tree, represent the quality requirements that the stakeholders consider to be more important. This initial set will be completed with the scenarios elicited from and prioritized in the brainstorming, which is the bottom-

up mechanism that complements the utility tree. This perspective is necessary as stakeholders do not usually think in terms of quality attributes but rather in terms of how the system will be used and the desired software qualities.

- The scenarios elicited from the brainstorming are classed, according to their purpose, in: use cases, which represent anticipated uses of the system; growth scenarios, which represent anticipated changes; and exploratory scenarios, or unanticipated stresses to the system. The basic references do not describe these classifications in the same way or state how they will be used later in the ATAM. The set of scenarios generated during brainstorming will be refined (grouping diverse scenarios related with the same action, refining their writing, etc.) and prioritized to select the scenarios that, by stakeholder agreement, must be satisfied by the architecture. This prioritization involves not only a vote to class the scenarios but also to assign the quality attribute(s) that each scenario affects most heavily.

- ABASs describe, sometimes implicitly, scenarios against which a certain quality attribute will be analyzed, on the basis of the architectural approach identified previously. An ABAS is determined by the general evaluation criterion (quality attribute) and the associated architectural style (for example, concurrent pipeline, synchronization, etc.). The stimuli for application and the responses to be output, as well as the architectural decisions associated with a given style, are identified in each ABAS. Therefore, it provides a subset of the quality attribute characterizations that are to be analyzed. Depending on the quality attribute in question, the response will be output by means of either a formula (for example, for the quality attribute performance in the Concurrent Pipeline ABAS) or developing scenarios (qualitative analysis), if there are no formulas applicable for outputting values for the responses (for example, for the quality attribute modifiability in the Data Indirection ABAS) [Klein 99]. On the basis of these scenarios, the data-gathering techniques needed to gather information on the responses described in the ABAS will be developed.

Therefore, the ATAM yardstick is the set of scenarios generated in a particular evaluation and used as a reference point to determine whether the elicited criteria are satisfied. This type of yardstick is not prescriptive in the sense that it does not provide fixed values (or ranges) for each criterion, and the values obtained from an evaluated architecture will later be compared against these values. In the ATAM, the values obtained for the criteria are judged in the context of a particular evaluation with a given architecture and with stakeholders who may vary (in number and type of stakeholder) from evaluation to evaluation.

3.3.2 Yardstick Elicitation

The elicitation of the yardstick can be analyzed from two viewpoints: during the development of the evaluation method, and in the execution of the ATAM. With regard to method development, the part of the yardstick that does not depend on the stakeholders, that is, the scenarios defined implicitly in ABASs, can be prepared. The development of ABASs is based on the identification of an architectural approach and the generation of a reasoning framework, "based on quality attribute-specific models, which exist in the various quality attribute communities" [Klein 99, p. 1]. Therefore, this part of the yardstick is generated by experts.

As far as method execution is concerned, the selection of the ABAS applicable to a given architecture (and therefore the elicitation of the subset of ABAS-dependent scenarios) is not

described in detail in the basic references. It is assumed that the evaluators will rely on the "criteria for choosing this ABAS," that have been defined for each ABAS; on their understanding of the software functionality (as the architectural approaches are identified after or simultaneously in step [1-3/2-4] "Present Architecture" shown in Figure 6 on page 23); and on their experience in relating part of the presented architecture directly to any ABAS. The final result is a set of architectural approaches identified in the evaluated architecture. This leads to the problem of combining the different approaches selected. Klein et al. describe how to combine like-attribute ABASs or ABASs for different attribute types [Klein 99, p. 16-23]. However, this combination is carried out for the purpose of synthesizing the information and getting the results of the evaluation (risks, non-risks, sensitivity points, and tradeoff points). Therefore, combinations of ABASs will be analyzed in Section 3.5. As yardsticks, scenarios are not combined but rather used independently.

With regard to the part of the yardstick dependent on the stakeholders (set of scenarios derived from the utility tree and brainstorming), it is prudent to describe these techniques in more detail, including examples and their interrelationship (for example, when the utility tree scenarios are used during an ATAM evaluation, how they are related with the scenarios elicited from the brainstorming, etc.) and to define and illustrate the use of the scenario classification elicited in the brainstorming (considering the published references, the most complete example is provided by Kazman et al. [Kazman 00, p. 33]).

3.4 Data-Gathering Techniques

3.4.1 Identification of Data-Gathering Techniques

The ATAM uses different techniques to obtain data about the architecture to be evaluated, such as presentations, group interviews, and brainstorming. However, taking into account the definition of a data-gathering technique (obtaining data to judge the target), not all these techniques can be classed as data-gathering techniques. For example, the ATAM uses presentations to know and delimit the system and to identify the architectural drivers, critical requirements, and architectural views, among other factors. Therefore, information is obtained, but it will be used to develop other evaluation components (target delimitation and understanding and criteria elicitation, if we considered the presentations). Another example is utility tree generation for identifying the quality attribute that will be considered in an evaluation [SEI 00, slide 49]. The characterizations associated with each quality attribute are the sources of questions to break down each quality attribute into specific subfactors (using the nomenclature of the ATAM) and to generate the scenarios of the leaves of the tree. These questions generated from the quality attribute characterizations and used to elicit the criteria and a part of the yardstick will not be considered as data-gathering techniques. But the questions generated from the quality attribute characterizations and used to obtain data to judge the architecture will be considered as data-gathering techniques. Therefore, the data-gathering techniques applied in the ATAM are the following:

- the set of questions developed and applied in step [1-6/2-7/2-9] "Analyze Architectural Approaches" (shown in Figure 6 on page 23) to obtain information about the target and apply the synthesis techniques

- scenario mapping in the evaluated architecture. In this mapping, evaluators can obtain information about architectural components and connectors. This information will also be used to analyze whether the architecture supports the quality attribute requirements.

- mathematical algorithms described in some ABASs to obtain a numerical value of some criteria. These algorithms are also data-gathering techniques that, considering their detail description based on mathematical concepts to calculate the response, have to be considered as defined and developed formally in the evaluation method.

From a conceptual point of view, data-gathering and synthesis techniques are applied in step [1-6/2-7/2-9] "Analyze Architectural Approaches" (shown in Figure 6 on page 23). The results of the evaluation—risks, non-risks, sensitivity points, and tradeoff points—are the main output of this step. In step [1-6/2-7], the three types of data-gathering techniques will be used; in step [2-9], only questions and mapping are used. Stakeholders play an active role in the generation of questions. Nevertheless, only those questions that the evaluators can generate and apply will be considered in this analysis of the ATAM, rather than those generated by the stakeholders. However, the discussion carried out during step [2-9] seems to be guided primarily by the stakeholders, while the mapping and trace are performed by the architect. In the description of step [2-9], it is not clear whether the evaluators will generate and apply questions to obtain more information.

The data-gathering and synthesis techniques of the ATAM are very closely interrelated. A deeper analysis of a concrete quality attribute, or the identification of a risk or sensitivity point can be carried out depending on how the architect answers the questions asked by evaluators. As those answers are obtained, evaluators identify and apply the synthesis techniques to determine the risks, non-risks, sensitivity points, and tradeoff points, and continue asking the architect more questions, as necessary. However, the independent identification and analysis of the data-gathering and synthesis techniques should facilitate the explicit and rigorous definition of these techniques.

3.4.2 Development of Data-Gathering Techniques

There are diverse sources to generate the questions to be applied, but these sources are not unified in the method documentation. The SEI ATAM presentation suggests some sources, but it is impossible to analyze the data-gathering techniques based on this reference alone because the definitions are excessively general [SEI 00]. The *ATAM Reference Guide* focuses on the questioner role, instead of stating when and how this evaluator will generate and apply the questions. Nevertheless, Barbacci et al. discuss the development of data-gathering techniques and identify a general outline that is applicable to all quality attributes and a classification of the diverse type of questions the evaluators will use. The general outline contains a set of questions grouped in three categories: "requirements," "for important services," and "for other services." The questions associated with each category are also classed taking into

account the following classification: screening, elicitation, and analysis questions. These questions will be applied in the ATAM "to collect and analyze information about current and future system drivers and architectural solutions" [Barbacci 00b]:

- *screening questions*: "used to narrow or focus the scope of the evaluation quickly. These questions are qualitative and not necessarily precise" [Barbacci 00b]. Based on this definition, it can be deduced that evaluators will apply these questions at the beginning of the ATAM evaluation to delimit the evaluation taking into account the quality attributes identified by the stakeholders. Therefore, this type of question (although necessary to delimit the target) is not included in the data-gathering techniques.

- *elicitation questions*: "used to gather information to be analyzed later. ... They collect information about decisions made, and the emphasis is on extracting quantifiable data ... These questions are guided by stimulus/response branches of the quality attribute [characterizations] ... or are guided by architecture mechanism branch of the quality attribute tables" [Barbacci 00b]. Therefore, elicitation and analysis questions are a part of the data-gathering techniques applied in the ATAM.

- *analysis questions*: "used to conduct analysis using attribute models and information collected by elicitation questions" [Barbacci 00b].

Kazman et al. mention this classification of questions and also present some examples of data-gathering techniques: for example, elicitation questions related to the quality attribute characterizations that help to ensure attribute coverage [Kazman 00, p. 10-12]. Also, Barbacci et al. state the order of applying these three types of questions. This order is related with two types of questioning that evaluators can apply: breadth-first questioning or depth-first questioning. The three basic references do not mention these strategies, and, in general, these types of questioning are not described in more detail (for example, to state if they are exclusive or complementary strategies). Also, the relationship between the three types of questions and the sources used by the evaluators to generate the questions are not established. Taking into account the relationship yardstick and data-gathering techniques and considering that a part of the yardstick is variable (stakeholder-dependent scenarios), it is not possible to generate a priori a standard list with all the questions to be applied. However, it could be possible for evaluators to generate the set of questions related with the architectural design-dependent scenarios (ABASs or at least the general structure to generate the questions, depending on the quality attribute considered). The evaluation team should have expert questioners in quality attribute areas identified during step [1-2/2-3] "Present Business Drivers" (according to the *ATAM Reference Guide*). These expert questioners will generate questions associated with the architectural approaches employed and verify that scenarios related to specific issues (if relevant to the architectural drivers at hand) will be elicited. However, the relationship between the types of questions (screening, elicitation, and analysis) and the questioner role is not established. Basic references do not describe if experts will develop only their own "set of questions" (based on the quality attribute of their expertise); or if they will develop only a specific type of questions (screening, etc.); or if a meeting with all team evaluation members will be carried out to create common data-gathering techniques with the purpose of assuring the development of a whole set of questions to analyze the architecture, independent of the

stakeholders questions. With regard to the stakeholders' questions, none of the basic references describe in detail the concrete participation of the stakeholders in step [1-6/2-7/2-9] "Analyze Architectural Approaches" (shown in Figure 6 on page 23). Nevertheless, considering the active participation of stakeholders in an ATAM evaluation, it is prudent that at least one evaluator verify the completeness of the analysis applied to each quality attribute; that is, determine if, for each quality attribute, there is sufficient data (derived from evaluators' and stakeholders' questions) to judge the evaluated architecture. Also, all the issues above described (relationships not identified, elements not described, etc.) should be addressed in order to know the specific approach used to develop the questions including, when possible, examples to illustrate the elaboration and use of questions.

Finally, with regard to the scenario mapping, the architect performs the mapping and answers the questions posed by evaluators and stakeholders during his or her explanation (according to the *ATAM Reference Guide*). This technique is very closely interrelated with the synthesis techniques because the architect's explanation and answers will be used to identify the risks, non-risks, sensitivity points, and tradeoffs points. Basic references do not describe this technique in detail, but Kazman et al. provide an example [Kazman 00, p.30] where the analysis of a utility tree scenario and the template to capture the information obtained during the mapping are shown. Nevertheless, this example is focused on the synthesis techniques, not on the data-gathering techniques. For example, the relationship between mapping and architectural views is not described. This relationship can be deduced only based on: the explicit references included in ABASs [Klein 99, p. 32] (although some ABASs do not include these references); the classification of the quality attribute stated by Bass et al. (runtime or development-time qualities); and by relating each quality attribute to the architectural views that allow us to represent these qualities. A more complete description of this technique would be an aid to understanding it.

3.5 Synthesis Techniques

3.5.1 Identification of Synthesis Techniques

Synthesis techniques are very closely interrelated with data-gathering techniques in the ATAM. Both types of techniques are applied in step [1-6/2-7/2-9] "Analyze Architectural Approaches" (shown in Figure 6 on page 23). Specifically, synthesis techniques correspond with the analysis realized to judge the architectural decisions and obtain the risks, non-risks, sensitivity points, tradeoff points, and sometimes recommendations: "any alternative architectures that should be considered" or "any architectural process issues that were uncovered" or "documentation practices that should be adopted evaluation" (according to the *ATAM Reference Guide*). In step [2-9], "the participants might identify risks, non-risks, sensitivity points, and tradeoff points;" so, evaluators as well as stakeholders participate in obtaining the results of the evaluation (according to the *ATAM Reference Guide*).

Nevertheless, the basic references do not describe in detail the techniques needed to obtain these final results. For example, the information obtained from questions, scenario mapping,

and ABAS algorithms is used in step [1-6/2-7/2-9] "Analyze Architectural Approaches" (shown in Figure 6 on page 23) to analyze the architectural decisions, taking into account the architectural approaches identified and, therefore, the set of selected ABASs and the reasoning associated with each of them. Also in this step, "the architect maps scenarios onto the appropriate architecture views, highlighting the components and connectors involved" (according to the *ATAM Reference Guide*). In the basic references, this step is described in such a general way that it is very difficult to analyze the synthesis techniques applied. As a result, it is prudent that method developers describe issues like the following in more detail: stakeholders' participation in obtaining the final results; the relationship of the analysis performed in step [1-6/2-7] with the mapping carried out in step [2-9]; how and when the scenarios obtained in the utility tree will be used; and, for example, how to relate the risks finally obtained with the quality attribute requirements in such a way that stakeholders will know which attributes are the sources of the key risks (or a great number of risks) and use this information to determine whether these qualities are strictly necessary or financially viable. Most of these issues were considered in the recent ATAM report that contains an "Architectural Approach Documentation Template" which shows the information related with the mapping and analysis of a scenario [Kazman 00].

3.5.2 Development of Synthesis Techniques

Considering that each ABAS includes an analysis section, in which the reasoning to apply for each architectural style is described, and a sample combination of similar or different attribute type ABASs, it can be deduced that a part of the synthesis techniques is developed [Klein 99]. Also, Klein et al. describe very general guidelines for identifying the tradeoffs (interactions across ABASs of different attribute types) [Klein 99, p. 21]. Nevertheless, it should be convenient to include a description, or a more detailed example, of the reasoning to apply in those ABASs for which analysis and reasoning are based only on scenarios.

With regard to step [1-6/2-7/2-9] "Analyze Architectural Approaches" and taking into account the basic references, we have to conclude that the synthesis techniques are not developed. However, these techniques are identified in the recent ATAM report [Kazman 00]. In this report we can find a template used to capture all the information elicited during the mapping and analysis of a scenario. This template and the associated example show the type of information we can obtain from an ATAM evaluation. So, evaluators will have an infrastructure (template) to apply in ATAM evaluations and an example to guide the evaluators to identify the risks, non-risks, sensitivity points, and tradeoff points. Nevertheless, to complete the description of the ATAM, it would be convenient (if it were possible) to explicitly describe guidelines applied to make judgments and obtain the final results of the evaluation. Also, it is necessary to consider that the importance of the stakeholders' questions posed as well as later analysis will depend completely on the set of prioritized scenarios and the stakeholders and their agreement "with captured summary before discussion moves on" (according to the *ATAM Reference Guide*). The stakeholders' participation in the analysis of the target, as well as in the definition of one part of the ATAM yardstick, highlight the importance of selecting the most appropriate stakeholders to be present in the evaluation.

3.6 Evaluation Process

In this section, the ATAM will be analyzed taking into account the three subprocesses (planning, examination, and decision making) and the main activities identified in Section 2.2.6, shown graphically in Figure 4 on page 15. The nomenclature subprocesses/activities/tasks will be used to refer to the evaluation process set out in the framework described in Section 2. The nomenclature phase/step/activities, defined in the *ATAM Reference Guide*, will be used to refer to the ATAM evaluation process.

The ATAM evaluation process definitions differ depending on the report addressee: evaluator or DO professional. If the document addresses the DO (or, generally, the stakeholders), it only describes the phases that call for participation and intensive cooperation by DO professionals. On the other hand, if the document addresses the evaluators, it offers a view of the entire evaluation process. All the steps shown in Figure 6 on page 23 represent the evaluation process described for the evaluators (according to the *ATAM Reference Guide*). The set of steps described for stakeholders fall within phases 1 and 2. In the SEI ATAM presentation, the steps of phases 1 and 2 are grouped according to a subclassification not mentioned in the *ATAM Reference Guide*: presentation, investigation and analysis, testing and out-briefing. During phase 2 and after the preparation step, steps [1-1] through [1-6] are repeated in the presence of the larger set of stakeholders, but steps [1-4/2-5] "Identify Architectural Approaches" and [1-5/2-6] "Generate Quality Attribute Utility Tree" are recapped and summarized for the larger audience.

Taking into account how the steps of the ATAM are carried out over time, phase 0 (Partnership and Preparation) tend to vary depending on factors such as: former evaluation team training; whether the evaluators have prior knowledge of the candidate system; or DO readiness to start negotiating the Statement of Work. According to the *ATAM Reference Guide*, the steps of the other phases must be performed in the order specified in Figure 6 on page 23, except step [3-1] "Produce the Final Report," which will be skipped if the DO does not require a written final report. The set of steps shown in Figure 6 will be considered in the next subsection to analyze the ATAM evaluation process against the subprocesses described in the framework.

3.6.1 Analysis of the ATAM Evaluation Process

Analyzing the ATAM evaluation process, as compared to the subprocesses and activities specified in Figure 4, it can be deduced that, generally, the ATAM includes the three subprocesses: planning, examination, and decision making. Table 9 shows the correspondence of the subprocesses and activities described in the framework with the phases and steps of the ATAM. The first three letters of the alphabet (subprocesses) and numbers (activities and tasks) are used to refer to the framework subprocesses and activities. To refer to the phases and steps of the ATAM, the pair [phase number–step number] is used, as shown in Figure 6.

Evaluation Framework Subprocesses and Activities	ATAM Phase-Step	ATAM Step Name
A. PLANNING		
A.1. Establish the evaluation goals		
A.1.1. Get to know and analyze the DO target	[0-2]	Description of the candidate system
A.1.1. Get to know and analyze the DO target	[0-5]	Forming the core evaluation team
A.1.2. Negotiate with the representatives of the DO	[0-1]	Present the ATAM
A.1.2. Negotiate with the representatives of the DO	[1-1 / 2-2]	Present the ATAM
A.1.3. Define the goals of the evaluation	[0-3]	Make a go/no-go decision
A.1.3. Define the goals of the evaluation	[0-4]	Negotiate the Statement of Work
A.2. Design the evaluation		
A.2.1. Target	[1-2 / 2-3]	Present business drivers
A.2.1. Target	[1-3 / 2-4]	Present architecture
A.2.2. Criteria	[1-2 / 2-3]	Present business drivers
A.2.2. Criteria	[1-3 / 2-4]	Present architecture
A.2.2. Criteria	[1-5 / 2-6]	Generate quality attribute utility tree
A.2.2. Criteria	[2-8]	Brainstorm and prioritize scenarios
A.2.3. Yardstick	[1-3 / 2-4]	Present architecture
A.2.3. Yardstick	[1-4 / 2-5]	Identify architectural approaches
A.2.3. Yardstick	[1-5 / 2-6]	Generate quality attribute utility tree
A.2.3. Yardstick	[1-6 / 2-7]	Analyze architectural approaches
A.2.3. Yardstick	[2-8]	Brainstorm and prioritize scenarios
A.2.4. Identification of the data-gathering techniques	[1-6 / 2-7 / 2-9]	Analyze architectural approaches
A.2.5. Identification of the synthesis techniques		
A.2.6. Evaluation process		
A.3 Analyze the target		
A.3.1. Request and analyze the general DO documentation	[0-7]	Prepare for phase 1
A.3.1. Request and analyze the general DO documentation	[2-1]	Prepare for phase 2
A.3.2. Set up the full evaluation team	[0-6]	Hold evaluation team kick-off meeting
A.3.2. Set up the full evaluation team	[2-1]	Prepare for phase 2
A.3.3. Identify the professionals involved	[1-2]	Present business drivers
A.3.4. Develop the data-gathering and synthesis techniques	[1-6 / 2-7]	Analyze architectural approaches
A.3.5. Develop/adapt the infrastructure		
A.4. Plan the evaluation	[0-7]	Prepare for phase 1
A.4. Plan the evaluation	[1-1 / 2-2]	Present the ATAM
A.4. Plan the evaluation	[2-1]	Prepare for phase 2

Table 9. Framework – ATAM Correspondence

Evaluation Framework Subprocesses and Activities		ATAM Phase-Step	ATAM Step Name
B. EXAMINATION			
	B.1. Apply the data-gathering techniques and obtain the data needed	[1-6 / 2-7 / 2-9]	Analyze architectural approaches
	B.2. Check the data gathered for completeness		
C. DECISION MAKING			
	C.1. Apply the synthesis techniques	[1-6 / 2-7 / 2-9]	Analyze architectural approaches
	C.2. Prepare the final report	[2-10]	Present results
		[3-1]	Produce the final report
	C.3. Present and submit the final report	[2-10]	Present results
		[3-1]	Produce the final report
	C.4. Complete the evaluation documentation	[3-2]	Hold post-mortem meeting
		[3-3]	Build portfolio and update artifact repositories

Table 9. Framework – ATAM Correspondence (cont'd.)

With regard to the repetition of some ATAM steps and taking into account the components of the evaluation and the characteristics of the method, we can deduce that some steps have to be repeated because there is a priori no specific delimitation of the target, no predetermined set of criteria for data gathering, and no yardstick as a reference point for the ATAM evaluation. As the evaluation cannot be run without these components, a set of processes have been planned, aimed primarily at getting the information required to roughly identify the possible target (delimitation), the set of criteria for evaluation, and the yardstick. This process is carried out together with three professionals (client, architect, and project manager) and later verified in the second phase when all the stakeholders are present. The following subsections present the analysis of the ATAM against the framework including the analysis of the order of the task performance and main differences detected.

3.6.1.1 Planning Subprocess

The planning subprocess encompasses the activities carried out from when contact is first made with the DO until all the evaluation components have been developed or adapted.

Establish the evaluation goals.

This activity corresponds to the first five steps of phase 0 of the ATAM, step 1 of phase 1, and step 2 of phase 2, shown in Figure 6 on page 23. The steps are not carried out in the same order as described in the framework, although the order of the steps may vary depending on several circumstances: standing agreement with the DO, knowledge of the system and needed iterations to achieve full understanding of the ATAM and candidate system (according to the *ATAM Reference Guide*). The main difference lies in the evaluators involved.

In the framework, it is assumed that the preliminary evaluation team (which will be a subset of the final team) will perform the activities. In the ATAM evaluation, the first four steps will be performed by EO technical staff or representatives of this organization; therefore, the core evaluation team is formed after assuring that the evaluation will occur. A further consequence is the need for the recently formed team to be informed about the target during step [0-6] "Hold Evaluation Team Kick-Off Meeting" (shown in Figure 6 on page 23). The evaluation can be canceled during either step [0-3] "Making a Go/No-Go Decision" or [0-4] "Negotiate the Statement of Work" as a result of not reaching an agreement.

With regard to step [1-1] "Present the ATAM," it was included because in this subprocess the evaluation method will be described to DO representatives. Therefore, basically the same actions are taken, although some differences (which should be convenient to address) are found:

- The *ATAM Reference Guide* does not include parameters to keep in mind when deciding whether to continue with the evaluation.

- In the *ATAM Reference Guide*, the description of step [0-5] "Form the Core Evaluation Team" states that "if there are changes to the candidate schedule then negotiate new schedule with customer; return to step [0-4]." Therefore, there is a cycle between these steps which is not reflected in step [0-4] "Negotiate the Statement of Work."

- The ATAM includes the team role definitions, specifying the responsibilities and desirable characteristics for each role. However, basic references do not include information about the necessary experience to play a role and the work done by and responsibility of two evaluation leaders.

- The basic references do not state whether the DO is always the sponsor organization. If it is possible for a company to request an ATAM evaluation of an architecture developed by a third organization, it would be necessary to identify with what organization this Statement of Work is negotiated and how the representatives of these companies participate in the evaluation.

Design the evaluation.

This activity corresponds to steps [1-2] to [1-6] and with steps [2-3] to [2-9] of the ATAM, as shown in Table 9. The main purpose of the framework activity "Design the evaluation" is the development or adaptation of the evaluation components taking into account the information about the target gathered in the preceding activity. Some activities in the framework have no corresponding ATAM step; synthesis techniques depend on the expert knowledge of evaluators, and there are no tasks in which the ATAM evaluation process, as described in the basic references, should be modified. With regard to the first three evaluation components, because there is a priori no specific delimitation of the target, no predetermined set of evaluation criteria, and no yardstick to apply, stakeholders must identify the target and select the criteria and yardstick to use in a particular evaluation. In general, the order in which the steps should be executed is the same as described in the framework, although in the ATAM the output of one step could be related to more than one evaluation component. Other issues that would

improve the evaluation process include the following (according to the *ATAM Reference Guide*):

- In step [1-2/2-3] "Present Business Drivers," with regard to stakeholders, the description of the last activity states that " ... team leader polls participants to ask for their lists of stakeholders roles, qualities, goals and constraints." Therefore, the participants in phase 1 (the client, architect, and project manager) will elaborate on the list of stakeholders that will take part in phase 2 of the ATAM evaluation. It is not clear whether evaluators consider other parameters in this selection, delimit the number of stakeholders if an excessive number is obtained, or specify individuals who must participate in the evaluation (if it is necessary).

- Steps [1-3/2-4] "Present Architecture" and [1-4/2-5] "Identify Architectural Approaches" are closely interrelated. The activities of both steps could be grouped in a unique step stating the explicit relationship and the order in which the activities should be executed. The current ATAM definition includes these two steps (instead of one) because the activities described are role oriented: there is no procedural description of the activities.

- The technique applied to prioritize quality attributes and subfactors included in the utility tree is not described in detail. For example, there is no description of how to select the concrete final set of criteria, where it will be used in later steps, or how to solve specific situations such as when most scenarios are identified as having the highest importance and maximum difficulty.

- The use of ABASs in steps [1-4/2-5] and [1-6/2-7/2-9] is not described in detail.

- In phase 2, the business drivers and architecture presentation will be carried out again. However, the ATAM description does not state whether these presentations must be exactly the same as those presented in phase 1, or whether the project manager or architect can modify the presentations. If modifications are accepted, the ATAM description should state whether the architect can modify the architectural views during the break between phases 1 and 2.

- The definition of current ATAM evaluation team roles should state that the evaluators must have a strong knowledge of the architectural styles.

Analyze the target.

This activity corresponds to steps [0-6] "Hold Evaluation Team Kick-Off Meeting" and [0-7] "Prepare for Phase 1," steps [1-2] "Present Business Drivers" and [1-6] "Analyze Architectural Approaches," and steps [2-1] "Prepare for Phase 2" and [2-7] "Analyze Architectural Approaches" of the ATAM, as shown in Table 9. However, these steps are not carried out in the same order as described in the framework. One of the main purposes of the "Analyze Target" activity is the development or adaptation of data-gathering and synthesis techniques. These techniques will be developed or adapted based on the documentation analysis provided by the DO. In the ATAM, these techniques will be developed as they are needed, during the same step in which they will be applied. Also, it is assumed that evaluators will use directly some or all of the templates included in the *ATAM Reference Guide*. Due to this, there is no step related with the development or adaptation of the infrastructure needed to perform the evaluation. Other issues that would improve the evaluation process include the following:

- In step [2-1] "Prepare for Phase 2," the core evaluation team will be completed "by adding questioners expert in quality attribute areas... "(according to the *ATAM Reference Guide*). Nevertheless, in this step activities to inform the new evaluators about the documentation, analysis, and results obtained in phase 1 are not included.

- The basic references do not state how to select new evaluation team members.

- The basic references do not mention the different types of questions that will be generated during the evaluation and whether the evaluation team can select the questions to apply in a particular evaluation from a pool of questions that were used in other evaluations.

- Synthesis techniques are not described in detail in the basic references.

Plan the evaluation.

This activity corresponds to steps [0-7] "Prepare for Phase 1," [1-1] "Present Business Drivers," [2-1] "Prepare for Phase 2," and [2-2] "Present the ATAM", as shown in Figure 6 on page 23. The main purpose of the "Plan Evaluation" framework activity is to plan all of the activities for the examination and decision-making subprocesses. The associated ATAM steps are performed in two different general periods: at the end of phase 0/beginning of phase 1, and at the beginning of phase 2. This execution order is derived directly from the need to plan two meetings with DO members.

3.6.1.2 Examination Subprocess

The examination subprocess encompasses the activities focused on the application of data-gathering techniques, after all of the evaluation components have been developed in the planning subprocess.

Apply the data-gathering techniques and gather the data needed.

This activity corresponds to step [1-6], [2-7], and [2-9] "Analyze Architectural Approaches" of the ATAM, as shown in Table 9. The framework differs from the ATAM. Conceptually, the activities (framework) and steps (ATAM) are carried out in different periods: in the framework, the data-gathering techniques will be applied after the evaluation components have been developed; in the ATAM, these techniques are applied in steps focused on the yardstick development. As the yardstick is developed, information about the evaluated architecture is gathered and a portion of the evaluation results is obtained. So, data-gathering and synthesis techniques are applied jointly. The following issues help to improve the evaluation process:

- Data-gathering techniques are not well defined in the basic references although Barbacci et al. discuss the strategy for generating the questions [Barbacci 00a, Barbacci 00b]. The *ATAM Reference Guide* and SEI ATAM presentation focus on other key concepts (such as scenarios, utility trees, etc.), but the relationship among these concepts and data-gathering techniques is not specified. The different types of questions that Barbacci et al. describe are mentioned in the last ATAM report [Kazman 00].
- Due to this, step [1-6/2-7/2-9] "Analyze Architectural Approaches" does not state whether evaluators analyze in advance the set of available questions to assure that

those questions sufficiently encompass the quality attributes to be analyzed, based on the identified architectural approaches.

- The ATAM evaluation process does not specify in detail when the data-gathering techniques should be applied.
- Data-gathering techniques are not related to the questioner role.
- In general, and considering the above comments, we can deduce that there is variability on the data-gathering techniques applied. This could cause variability in the final results of the evaluation due to the dependency on both the evaluators (and their knowledge, skills, and experience) and the stakeholders. In the ATAM evaluation, the variability dependent on the stakeholders will be a constant factor, but the variability dependent on the evaluators can be minimized by describing explicitly the evaluation components (utility tree, ABAS, questions, etc.) and their interrelationships.

Check the data gathered for completeness.

This framework activity does not correspond with the ATAM steps or activities, as they are described in the basic references. Nevertheless, due to the active participation of the stakeholders in the evaluation (asking questions to the architect, etc.), it would be prudent for an evaluator to analyze the information gathered to determine whether there is sufficient data about the scenarios to be applied during the evaluation to judge the evaluated architecture.

3.6.1.3 Decision-Making Subprocess

The decision-making subprocess encompasses the activities focused on applying the synthesis techniques to obtain the final results of the evaluation, and in activities related to preparing for the final written report and the end of the evaluation.

Apply the synthesis techniques.

This activity corresponds to steps [1-6], [2-7], and [2-9] "Analyze Architectural Approaches" of the ATAM, as shown in Table 9. The timeframe for applying the synthesis techniques differs in the framework and the ATAM. In the framework, these techniques are applied after gathering all the information needed to judge the target. While in the ATAM, they are applied in steps focused on the yardstick development and, as the yardstick is developed and evaluators gather information about the evaluated architecture (application of the data-gathering techniques), they obtain the evaluation results. Nevertheless, due to the general description of the framework activities and the ATAM steps included in the basic references, it is very difficult to analyze the synthesis techniques in detail; it is unclear when and how to develop them or when and how to apply them. This is one of the key improvements to the method. Nevertheless, in the last ATAM report this improvement was partially undertaken. Method developers included a template for capturing the information obtained after applying the synthesis techniques.

Prepare the final report.

This activity corresponds to steps [2-10] "Present Results" and step [2-1] "Prepare for Phase 2" of the ATAM, as shown in Table 9. Differences in the execution order are not detected,

because these steps will be performed after the final evaluation results are obtained. In the ATAM, the findings will be presented in a presentation, and, if it was requested in the Statement of Work, a final report will be produced in phase 3. The template for the final report is included in the ATAM Reference Guide.

Present and submit the final report.

This activity corresponds to steps [2-10] "Present Results" and [3-1] "Produce the Final Report" of the ATAM, shown in Table 9. Differences in the execution order exist but are not significant, because in the ATAM the results presentation is carried out at the end of phase 2.

Complete the evaluation documentation.

This activity corresponds to steps [3-2] "Hold the Post-Mortem Meeting" and [3-3] "Build Portfolio and Update Artifact Repositories" of the ATAM, as shown in Table 9. In the framework, this activity involves collecting all the information and documentation used and/or generated with the purpose of analyzing and improving the evaluation method, and maintaining the documentation so that these results can later be compared to the outputs of future evaluations. The ATAM also includes tasks with the same purpose. However, the information to be included in the evaluation portfolio does not encompass all the material used/generated, but rather includes the copy of presentation results, the final report, the participant and team evaluations, the process observer's report, and the long-term benefit survey.

4 Conclusions and Recommendations

As a summary, Table 10 shows the results of the matching between the ATAM and the framework. The main characteristics of the ATAM that differentiate this method from other types of evaluation are the following:

- an evaluation team that includes experts in each quality attribute to be analyzed in a specific evaluation

- There is not a priori standard set of criteria or yardstick use when performing an ATAM evaluation; instead, each application of the method will elicit the criteria and yardstick that are appropriate for it.

- The yardstick is composed of a set of scenarios generated in each evaluation. Current ATAM evaluations have a great number of stakeholder-dependent scenarios.

- a strong collaboration of the stakeholders in the evaluation process, to identify and develop the components as well as to participate in the application of the data-gathering and synthesis techniques

Evaluation Components	ATAM
Target	Architecture
Criteria	General criteria: quality attributes Specific criteria: architectural decisions, responses
Yardstick	Set of scenarios generated (implicitly or explicitly) in a particular evaluation
Data-gathering techniques	Questions used to obtain information about the architectural decisions and the responses Mapping scenarios onto the architecture Algorithms described in some ABASs
Synthesis techniques	Analysis performed in step [1-6/2-7/2-9] to obtain the results of the ATAM evaluation: risks, non-risks, sensitivity points, and tradeoff points
Evaluation process	Steps and tasks described in the ATAM documentation

Table 10. Identification of Evaluation Components for the ATAM

Based on these characteristics, the ATAM can be classed as an eclectic evaluation method, which presents characteristics of an expertise-oriented evaluation (performed by experts), a management-oriented evaluation (taking into account that stakeholders have a strong participation and that the evaluation results will help them make decisions about the quality attributes and the architecture), and an objective-oriented evaluation (based on the use of applied

scenarios as a reference point to analyze whether the architecture satisfies their quality requirements).

Analyzing the evaluation process of the ATAM during the last two years, a significant change has taken place: the evaluation process described by Kazman et al. is focused on the scenario concept while the evaluation process defined in the current basic references introduces the concept of architectural style, and, in general, the ABAS [Kazman 99]. The handbook of ABASs will contain not only the architectural styles taken as reference points to evaluate and develop an architecture, but also the expert knowledge of the evaluators to judge the architecture and identify the risks, non-risks, sensitivity points, and tradeoff points, and the alternatives that can be used to improve the architecture. Once the ABASs handbook is available, the evaluation will be based on architectural styles described explicitly and the judgement associated with each style. In that moment, ABASs will be the main component of the ATAM evaluation's yardstick, pushing scenarios into the background.

Nevertheless, until then, the scenarios will have a relevant role in the ATAM evaluation, and therefore the participation of the stakeholders is critical: the ATAM should control the roles of stakeholders who will participate in the evaluation and identify any required roles. This is one of the key possible improvements to the current ATAM description. The next issues summarize the proposals stated in Section 3 and/or include new general suggestions:

- The explicit description of the target (system architecture, software architecture, or software system's architecture) will be an aid to delimit specifically the evaluation method. This will have an impact on the potential stakeholders because they will know exactly what will be evaluated and, as result, when to require an ATAM evaluation.

- Due to the fact that the ATAM uses terminology closely related with OO systems (e.g., use cases) and that the Architecture-Based Design method states that use cases capture functional requirements (in addition to quality requirements), it would be convenient to describe explicitly whether the ATAM evaluates functional requirements in some way or focuses only on quality requirements [Bachmann 00, p. 7]. Applying multiple meanings to one term could result in misunderstandings about the scope of an ATAM evaluation.

- The method description should be analyzed taking into account the potential uses of the ATAM evaluation: for the selection of alternative candidate architectures, the enhancement of an evaluated architecture, and so on. This will provide a more accurate method description because the number of organizations to be involved in the evaluation and the roles played by the different stakeholders from all the companies during an ATAM evaluation will be specified explicitly.

- A more detailed description of the evaluation team, the requirements needed to be an ATAM evaluator, and the optimum number of evaluators in each team will help select the appropriate team for an evaluation and train and prepare new team members.

- With regard to evaluation criteria, diffusing the basic concepts of the ATAM (quality attribute characterizations, architectural styles, ABASs, etc.) will help make the communication process among stakeholders, software system developers, and evaluation team members easier, because it will provide a common view (in meaning as well as in nomenclature) of the quality attributes and the subfactors associated with each one.

- If the method considers the interrelations among evaluation components, the evaluation process will be refined, because the tradeoffs between components will be known. The explicit identification of the evaluation components and their relationship would help identify when and how to apply/use each element. For example, knowing the relationship between evaluation criteria, the yardstick, and data-gathering techniques will allow the evaluators to develop questions to be applied based on the quality attributes (and subfactors) considered, the set of scenarios (and the environment described in each one), and the questioning schema described by Barbacci et al. [Barbacci 00b]. Knowing these relationships will improve the evaluation process, because the evaluation method developers will be able to identify which activities are needed to develop each evaluation component, when to apply them, and how a change in one component may affect other components.

- The development of basic architectural styles (ABASs) described above will help
  - developers produce an architecture
  - evaluators and stakeholders identify the styles used in an evaluated architecture and therefore create a yardstick based on architectural styles, not only on scenarios
  - evaluators because they will have the guidelines needed to judge the evaluated architecture(s) based on the selected ABASs and to propose alternatives to enhance the quality of the evaluated architecture

- Finally, with regard to the evaluation process, it would be defined more rigorously if each step were described with more detail, including a thorough description of each activity, the relationships among activities/steps, and the evaluation components elicited/developed/applied in each step. In addition, it will be helpful if the evaluation process includes a set of tasks to be performed by the DO to prepare it for the ATAM evaluation. In this way, potential DOs would know (and compare) the effort they have to support and the advantages provided by the results of an ATAM evaluation.

References/Bibliography

[SEI 00]　　　　"The Architecture Tradeoff Analysis Method" presentation. Available WWW: < URL: http://www.sei.cmu.edu/ata/ATAM/index.htm> (2000).

[Bachmann 00]　　Bachmann, F.; Bass, L.; Chastek, G.; Donohoe, P.; & Peruzzi, F. *The Architecture-Based Design Method* (CMU/SEI-2000-TR-001 ADA375851). Pittsburgh, Pa.: Software Engineering Institute, Carnegie Mellon University, 2000. Available WWW: < URL: http://www.sei.cmu.edu/publications/documents/00.reports/00tr001.html)> (2000).

[Barbacci 00a]　　Barbacci, M. "Quality Attribute Workshops." *news@sei interactive 3*, 2 (Spring 2000). Available WWW: <URL: http://interactive.sei.cmu.edu/news@sei/columns/the_architect/2000/spring/the_architect.htm> (2000).

[Barbacci 00b]　　Barbacci, M.; Ellison, R.; Weinstock, C.; & Wood, W. *Quality Attribute Workshop Participants Handbook* (CMU/SEI-2000-SR-001). Pittsburgh, Pa.: Software Engineering Institute, Carnegie Mellon University, 2000. Available WWW: <URL: <http://www.sei.cmu.edu/activities/ata/Workshop_Book.pdf> (2000).

[Barbacci 95]　　Barbacci, M.; Klein, M.; Longstaff, T.; & Weinstock, C. *Quality Attributes* (CMU/SEI-95-TR-021 ADA307888). Pittsburgh, Pa.: Software Engineering Institute, Carnegie Mellon University, 1995. Available WWW: <URL: http://www.sei.cmu.edu/publications/documents/95.reports/95.tr.021.html> (1995).

[Barbacci 99]　　Barbacci, M. "Analyzing Quality Attributes." *news@sei interactive 2*, 1 (March 1999). Available WWW: <URL: http://interactive.sei.cmu.edu/columns/the_architect/1999/march/architect.mar99.pdf> (1999).

[Bass 98]　　　　Bass, L.; Clements, P.; & Kazman, R. *Software Architecture in Practice*. SEI Series in Software Engineering. Reading, Ma: Addison-Wesley, 1998.

[Boehm 00]　　　Boehm, B. "Unifying Software Engineering and Systems Engineering." *IEEE Computer 33*, 3 (March 2000): 114-116.

[CMMI 99] "Transition to CMMI[SM] Models" presentation. *Software Technology Conference (STC '99).* Salt Lake City, Utah, May 6, 1999. Available WWW: <URL: http://www.sei.cmu.edu/cmm/cmmi/stc99.presen/index.htm> (1999).

[Jones 99] Jones, L. & Kazman, R. "Software Architecture Evaluation in the DoD Systems Acquisition Context." *news@sei interactive 4,* 2 (December, 1999). Available WWW: <URL: http://interactive.sei.cmu.edu/Columns/The_Architect/1999/December /Architect.dec99.pdf> (1999).

[Kazman 00] Kazman, R.; Klein, M.; & Clements, P. *ATAM[SM]: Method for Architecture Evaluation* (CMU/SEI-2000-TR-004). Pittsburgh, Pa.: Software Engineering Institute, Carnegie Mellon University, 2000. Available WWW: <URL: http://www.sei.cmu.edu/publications/documents/00.reports/00tr004.ht ml> (2000).

[Kazman 99] Kazman, R.; Barbacci, M.; Klein, M.; Carrière, J.; & Woods, S. "Experience with Performing Architecture Tradeoff Analysis," 54-63. *Proceedings of ICSE'99.* Los Angeles, California, May 16-22, 1999. Available WWW: <URL: http://www.sei.cmu.edu/ata/icse99.pdf> (1999).

[Klein 99] Klein, M. & Kazman, R. *Attribute-Based Architectural Styles* (CMU/SEI-99-TR-022 ADA371802). Pittsburgh, Pa.: Software Engineering Institute, Carnegie Mellon University, 1999. Available WWW: <URL: http://www.sei.cmu.edu/publications/documents/99.reports/ 99tr022/99tr022abstract.html> (1999).

[Scriven 00] Scriven, M. Evaluation Core Courses Notes. Claremont Graduate University. Available WWW: <URL: http://eval.cgu.edu/lectures/lecturen.htm> (2000).

[Scriven 91] Scriven, M. *Evaluation Thesaurus.* Newbury Park, CA: Sage Publications, 1991.

[Shadish 91] Shadish, W.; Cook, T.; & Leviton, L. *Foundations of Program Evaluation. Theories of Practice.* Newbury Park, CA: Sage Publications, 1993.

[Stufflebeam 84] Stufflebeam, D. & Shinkfield, A. *Systematic Evaluation: A Self-Instructional Guide to Theory and Practice.* Boston, MA: Kluwer-Nijhoff Publishing, 1984.

[SM] CMMI and ATAM are service marks of Carnegie Mellon University.

[Taylor 96]   Taylor-Powell, E. & Steele, S. *Collecting Evaluation Data: An Overview of Sources and Methods* (Report: G3658-4). Program Development and Evaluation, University of Wisconsin-Extension, 1996. Available WWW: <URL: http://cf.uwex.edu/ces/pubs/pdf/ G3658_4.pdf> (1996).

[Worthen 97]   Worthen, B.; Sanders, J.; & Fitzpatrick, J. *Program Evaluation. Alternative Approaches and Practical Guidelines*. New York, N.Y.: Addison Wesley Longman, 1997.

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (LEAVE BLANK)	2. REPORT DATE September 2000	3. REPORT TYPE AND DATES COVERED Final

4. TITLE AND SUBTITLE Evaluation Theory Perspective of the Architecture Tradeoff Analysis Method (ATAM)	5. FUNDING NUMBERS F19628-00-C-0003

6. AUTHOR(S)
Marta Lopez

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Software Engineering Institute Carnegie Mellon University Pittsburgh, PA 15213	8. PERFORMING ORGANIZATION REPORT NUMBER CMU/SEI-2000-TR-012

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) HQ ESC/XPK 5 Eglin Street Hanscom AFB, MA 01731-2116	10. SPONSORING/MONITORING AGENCY REPORT NUMBER ESC-TR-2000-012

11. SUPPLEMENTARY NOTES

12.A DISTRIBUTION/AVAILABILITY STATEMENT Unclassified/Unlimited, DTIC, NTIS	12.B DISTRIBUTION CODE

13. ABSTRACT (MAXIMUM 200 WORDS)

Evaluation is a key analytical process in all disciplines and intellectual and practical endeavors. Also it is a key process in the software engineering field in which it is possible to apply different types of evaluation methods. The study of diverse evaluation methods performed in software and non-software disciplines and theoretical concepts could provide knowledge of the complexity and ubiquity of this important process. This study was the basis to obtain a set of basic evaluation components. They constitute a framework which can be used to developed a new evaluation method or review an existing one with the purpose of improving the development of the method being analyzed. In particular, this framework had been applied to review the Architecture Tradeoff Analysis Method (ATAM) by means of the identification of the evaluation components and the analysis of their development or elicitation. In this paper, the target, evaluation criteria, yardstick, data-gathering techniques, synthesis techniques and evaluation process of ATAM have been identified and analyzed. The most relevant conclusions are the role of stakeholders and the significance of attribute-based architectural styles (ABASs) in an ATAM evaluation.

14. SUBJECT TERMS software architecture, architecture analysis, ATAM, quality attributes, evaluation theory	15. NUMBER OF PAGES 62

16. PRICE CODE

7. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UL